

Université de Montréal

# **Étude des interactions supramoléculaires par modélisation moléculaire**

par

Marc-André Dubois

Département de Chimie

Faculté des arts et des sciences

Mémoire présenté à la Faculté des études supérieures et postdoctorales  
en vue de l'obtention du grade de Maître ès sciences (M.Sc.)  
en chimie

novembre 2012

© Marc-André Dubois, 2012

Université de Montréal  
Faculté des études supérieures et postdoctorales

Ce mémoire intitulé :

Étude des interactions supramoléculaires par modélisation moléculaire

Présenté par :

Marc-André Dubois

a été évalué par un jury composé des personnes suivantes :

Radu Ion Iftimie, président-rapporteur  
Andreea R. Schmitzer, directrice de recherche  
James D. Wuest, codirecteur  
Alexis Vallée-Bélisle, membre du jury

## Résumé

L'avancée des infrastructures informatiques a permis l'émergence de la modélisation moléculaire. À cet effet, une multitude de modèles mathématiques sont aujourd'hui disponibles pour simuler différents systèmes chimiques.

À l'aide de la modélisation moléculaire, différents types d'interactions chimiques ont été observés. À partir des systèmes les plus simples permettant l'utilisation de modèles quantiques rigoureux, une série d'approximations a été considérée pour rendre envisageable la simulation de systèmes moléculaires de plus en plus complexes.

En premier lieu, la théorie de la fonctionnelle de densité dépendante du temps a été utilisée pour simuler les énergies d'excitation de molécules photoactives. De manière similaire, la DFT indépendante du temps a permis la simulation du pont hydrogène intramoléculaire de structures analogues au 1,3,5-triazapentadiène et la rationalisation de la stabilité des états de transition.

Par la suite, la dynamique moléculaire et la mécanique moléculaire ont permis de simuler les interactions d'un trimère d'acide cholique et d'un pyrène dans différents solvants. Cette même méthodologie a été utilisée pour simuler les interactions d'un rotaxane-parapluie à l'interface d'un système biphasique.

Finalement, l'arrimage moléculaire et les fonctions de score ont été utilisés pour simuler les interactions intermoléculaires entre une protéine et des milliers de candidats moléculaires. Les résultats ont permis de mettre en place une stratégie de développement d'un nouvel inhibiteur enzymatique.

**Mots-clés :** Modélisation moléculaire, DFT, GROMACS, GAFF, FITTED, GAUSSIAN, 1,3,5-triazapentadiène, rotaxane, acide cholique, pyrène

## **Abstract**

The evolution of computer systems has led to the emergence of molecular modeling. To this end, a variety of mathematical models are now available to simulate various chemical systems.

Using molecular modeling, different types of chemical interactions were observed. From the simplest systems allowing the use of rigorous quantum models, a series of approximations were considered in order to make possible the simulation of increasingly complex molecular systems.

First, time-dependent density functional theory has been used to simulate the excitation energies of photoactive molecules. Similarly, time-independent DFT has enabled the simulation of intramolecular hydrogen bonding in the 1,3,5-triazapentadiene system and the rationalization of the stability of the transition states.

Subsequently, molecular dynamics and molecular mechanics were used to simulate the interactions of a trimer of cholic acid with a pyrene in different solvents. This methodology was then used to simulate the interactions of an umbrella-rotaxane at the interface of a biphasic system.

Finally, molecular docking and the concept of scoring functions were used to simulate the intermolecular interactions between a protein molecule and thousands of potential ligands. The results were then used to create a strategy for the development of a new enzyme inhibitor.



**Keywords** : Molecular modeling, DFT, GROMACS, GAFF, FITTED, GAUSSIAN, 1,3,5-triazapentadiene, rotaxane, cholic acid, pyrene

# Table des matières

<b>Chapitre 1: Introduction</b>	<b>1</b>
<b>1.1. Les mathématiques et la connaissance</b>	<b>1</b>
<b>1.2. L'évolution des modèles mathématiques en chimie</b>	<b>3</b>
1.2.1. La découverte de l'électron et le modèle <i>Plum Pudding</i>	3
1.2.2. Le modèle Rutherford-Bohr	3
1.2.3. Le comportement ondulatoire de la matière	5
<b>1.3. L'équation de Schrödinger</b>	<b>6</b>
<b>1.4. L'approximation de Born-Oppenheimer</b>	<b>7</b>
<b>1.5. Les calculs de structure électronique</b>	<b>8</b>
1.5.1. La théorie Hartree-Fock	8
1.5.2. Les méthodes Hartree-Fock semi-empiriques	9
1.5.3. La théorie de la fonctionnelle de densité	11
1.5.4. Le modèle Kohn-Sham	12
1.5.5. Approximation locale de la densité	13
1.5.6. Approximation des gradients généralisée	14
1.5.7. Méthodes hybrides	14
<b>1.6. La mécanique moléculaire</b>	<b>15</b>
1.6.1. Potentiel de liaison	16
1.6.2. Potentiel d'angle	17
1.6.3. Potentiel d'angle dièdre	18
1.6.4. Potentiel électrostatique	19
1.6.5. Potentiel de Lennard-Jones	20
1.6.6. Les champs de force	21
<b>Chapitre 2: Utilisation de la théorie de la fonctionnelle de la densité dans l'étude des interactions intermoléculaires</b>	<b>22</b>
<b>2.1. Choix de la DFT</b>	<b>25</b>
2.1.1. B3LYP	26
2.1.2. PBE0	27

2.1.1.	Fonctions de base .....	28
2.1.2.	TD-DFT .....	31
2.1.3.	Choix de GAUSSIAN .....	31
<b>2.2.</b>	<b>Méthodologie .....</b>	<b>31</b>
2.2.1.	Construction des structures .....	32
2.2.2.	Détermination des états de transition .....	32
2.2.3.	Observation des orbitales naturelles.....	33
2.2.4.	Détermination du spectre d'absorption par TD-DFT .....	34
<b>2.3.</b>	<b>Détermination des énergies d'excitation de molécules photoactives .....</b>	<b>36</b>
2.3.1.	Sommaire des résultats expérimentaux disponibles.....	39
2.3.2.	Résultats computationnels.....	40
2.3.3.	Conclusions .....	44
<b>2.4.</b>	<b>Observation des interactions intramoléculaires par l'étude des états de transition de molécules analogues au 1,3,5-triazapentadiène .....</b>	<b>44</b>
2.4.1.	Résultats computationnels.....	46
2.4.2.	Limites et conclusions.....	55
<b>Chapitre 3:</b>	<b>Utilisation de la dynamique moléculaire dans l'étude des interactions supramoléculaires .....</b>	<b>57</b>
<b>3.1.</b>	<b>Mécanique statistique.....</b>	<b>59</b>
<b>3.2.</b>	<b>Choix du champ de force .....</b>	<b>61</b>
3.2.1.	MMFF94.....	62
3.2.2.	CGENFF (CHARMM) .....	63
3.2.3.	OPLS .....	64
3.2.4.	GAFF (AMBER) .....	65
<b>3.3.</b>	<b>Choix de la dynamique moléculaire, de GROMACS et d'Antechamber .....</b>	<b>66</b>
<b>3.4.</b>	<b>Méthodologie .....</b>	<b>67</b>
3.4.1.	Construction des structures et calcul des charges atomiques.....	68
3.4.2.	Solvatation.....	68
3.4.3.	Traitement des interactions non liantes .....	69
3.4.4.	Minimisation d'énergie .....	70

3.4.5. Algorithme d'intégration <i>Leap-Frog</i> .....	70
3.4.6. Thermostat de Berendsen .....	71
3.4.7. Barostat de Parrinello-Rahman .....	72
3.4.8. Phase d'équilibration .....	74
3.4.9. Phase de production .....	75
<b>3.5. Simulation de l'encapsulation de pyrène par un trimère d'un dérivé d'acide cholique</b> .....	<b>75</b>
3.5.1. Sommaire des résultats expérimentaux disponibles .....	76
3.5.2. Résultats computationnels .....	78
3.5.3. Conclusions .....	84
<b>3.6. Simulation des interactions d'un rotaxane-parapluie à l'interface d'une séparation de phase entre l'eau et le dichlorométhane</b> .....	<b>85</b>
3.6.1. Sommaire des résultats expérimentaux disponibles .....	86
3.6.2. Résultats computationnels .....	87
3.6.3. Limites et conclusions .....	98
<b>3.7. Conclusions générales</b> .....	<b>98</b>
<b>Chapitre 4: Utilisation de l'arrimage moléculaire comme stratégie de développement d'un nouvel inhibiteur enzymatique</b> .....	<b>100</b>
4.1. Algorithmes génétiques et arrimage moléculaire .....	101
4.2. Contexte de l'étude .....	105
4.3. Méthodologie .....	106
4.3.1. Structure de la protéine .....	107
4.3.2. Fonction de score et FITTED .....	107
4.3.3. Choix des molécules cibles pour le criblage virtuel .....	111
4.1. Sommaire des résultats expérimentaux .....	111
4.2. Validation du modèle .....	112
4.3. Arrimage moléculaire .....	113
4.4. Criblage virtuel .....	115
4.5. Limites et conclusions .....	123
<b>Chapitre 5: Conclusion</b> .....	<b>125</b>

## Liste des figures

Figure 1.1 – Étude par éléments finis d’une pièce mécanique.....	2
Figure 1.2 – Oscillateur harmonique décrivant l’énergie d’étirement d’un lien covalent ...	17
Figure 1.3 – Oscillateur harmonique décrivant l’énergie de déformation de l’angle entre deux liens covalents .....	18
Figure 1.4 – Fluctuation de l’énergie en fonction de l’angle dièdre H-N-C-O d’une liaison peptidique.....	19
Figure 1.5 – Effet de la distance sur le potentiel électrostatique entre deux atomes partiellement chargés positivement.....	20
Figure 1.6 – Potentiel de Lennard-Jones entre deux atomes neutres .....	21
Figure 2.1 – Illustration d’un trou de Coulomb et de Fermi autour d’un électron.....	23
Figure 2.2 – Orbitales canoniques de valence du méthane .....	33
Figure 2.3 – Orbitales naturelles de valence du méthane.....	34
Figure 2.4 – Schéma d’une cellule photovoltaïque organique.....	36
Figure 2.5 –Fonctionnement de la couche active.....	37
Figure 2.6 – Molécules à l’étude.....	39
Figure 2.7 – Géométrie optimisée de la structure <b>2.1</b> et <b>2.2</b> .....	40
Figure 2.8 – Encombrement stérique entre le cœur et les extrémités du composé <b>2.1</b> .....	41
Figure 2.9 – Géométrie optimisée de la structure <b>2.3</b> et <b>2.4</b> .....	41
Figure 2.10 – Orbitales HOMO et LUMO des composés <b>2.1</b> à <b>2.4</b> .....	42
Figure 2.11 – Fragments à l’étude.....	45
Figure 2.12 – Motif de reconnaissance bis-diaminotriazine <b>2.13</b> .....	46
Figure 2.13 – Orbitale naturelle de la paire libre des molécules <b>2.5</b> et <b>2.12</b> .....	48
Figure 2.14 – Relation entre la variation de distance proton-accepteur ( $\Delta d_{H-Acc}$ ) et la différence d’énergie de l’état de transition ( $\Delta \Delta E_{Tr}$ ) .....	49
Figure 2.15 – Angles $\alpha$ , $\beta$ , $\gamma$ et $\delta$ pour la molécule <b>2.5</b> .....	50
Figure 2.16 – Relation entre la variation de l’angle combiné ( $\Delta \angle_{comb}$ ) et la différence d’énergie de l’état de transition ( $\Delta \Delta E_{Tr}$ ) .....	51

Figure 2.17 – Orbitales naturelles $\pi$ du composé <b>2.5</b> et <b>2.9</b> .....	53
Figure 2.18 – Orbitales naturelles $\pi$ du composé <b>2.7</b> et <b>2.5</b> .....	54
Figure 2.19 – Orbitales naturelles $\pi$ du composé <b>2.5</b> et <b>2.6</b> .....	55
Figure 3.2 – Illustration de la tyrosine et du phénol .....	63
Figure 3.3 – Représentation d’une boîte de solvant par la méthode PBC.....	68
Figure 3.4 – Effet du thermostat de Berendsen sur la température d’un système aqueux ...	72
Figure 3.5 – Effet du barostat de Parinello-Rahman sur la masse volumique d’un système aqueux .....	74
Figure 3.6 – Structure de l'acide cholique.....	76
Figure 3.7 – Trimère carbamide de l'acide cholique .....	77
Figure 3.8 – Structure des sondes utilisées pour les analyses spectroscopiques.....	77
Figure 3.9 – Absorbance du composé <b>3.2</b> en absence et en présence du trimère <b>3.1</b> et fluorescence du composé <b>3.3</b> en absence et en présence du trimère <b>3.1</b> .....	78
Figure 3.10 – Valeurs $\delta$ entre la sonde <b>3.3</b> et le trimère <b>3.1</b> dans l'eau et dans le THF .....	80
Figure 3.11 – Conformation finale du pyrène <b>3.3</b> et du trimère <b>3.1</b> dans l'eau.....	81
Figure 3.12 – Conformation finale du pyrène <b>3.3</b> et du trimère <b>3.1</b> dans le THF.....	82
Figure 3.13 – Valeur $\delta$ entre la sonde <b>3.4</b> et le trimère <b>3.1</b> dans l'eau et dans le THF.....	82
Figure 3.14 – Conformation finale du pyrène <b>3.4</b> et du trimère <b>3.1</b> dans l'eau.....	83
Figure 3.15 – Conformation finale du pyrène <b>3.4</b> et du trimère <b>3.1</b> dans le THF.....	84
Figure 3.16 – Schéma d'un rotaxane .....	85
Figure 3.17 – Illustration de l’axe <b>3.5</b> , de la roue <b>3.6</b> et de l’assemblage supramoléculaire du rotaxane-parapluie <b>3.5 + 3.6</b> .....	86
Figure 3.18 – Valeurs $\delta$ entre les bras et le cœur polaire du rotaxane <b>3.5 + 3.6</b> .....	87
Figure 3.19 – Conformation du rotaxane-parapluie <b>3.5 + 3.6</b> dans le dichlorométhane ....	88
Figure 3.20 – Conformation du rotaxane-parapluie <b>3.5 + 3.6</b> dans l'eau.....	89
Figure 3.21 – Conformations initiales des simulations de séparation de phase.....	90
Figure 3.22 – Trajectoire du rotaxane <b>3.5 + 3.6</b> initialement placé dans la partie lipophile du système biphasique .....	91

Figure 3.23 – Trajectoire du rotaxane <b>3.5 + 3.6</b> initialement placé dans la partie hydrophile du système biphasique .....	92
Figure 3.24 – Rapprochement d'un des bras du rotaxane <b>3.5 + 3.6</b> à l'interface .....	93
Figure 3.25 – Illustration du cœur polaire et de la portion hydrophile de l'acide cholique à l'interface .....	94
Figure 3.26 – Évolution de la distance entre le bras et le cœur polaire du rotaxane <b>3.5 + 3.6</b> .....	95
Figure 3.27 – Rapprochement du second bras à l'interface .....	96
Figure 3.28 – Passage du cœur dans la phase lipophile .....	97
Figure 4.1 – Surface de potentiel pour deux degrés de liberté.....	101
Figure 4.2 – Schéma itératif typique d'un algorithme génétique.....	102
Figure 4.3 – Dérivé de benzoporphyrine <b>4.1</b> .....	106
Figure 4.4 – Pourcentage de succès de différentes expériences d'arrimage avec plusieurs logiciels .....	109
Figure 4.5 – Comparaison des constantes de liaison de l'anticorps avec la protéine en présence des différents candidats avec les scores obtenus par l'arrimage.....	112
Figure 4.6 – Zone d'arrimage de la benzoporphyrine <b>4.1</b> dans la protéine .....	114
Figure 4.7 – Inverse des valeurs $\gamma$ pour tous les candidats .....	116
Figure 4.8 – Valeurs $\gamma$ pour tous les candidats.....	117
Figure 4.9 – Arrimage des meilleurs candidats.....	118
Figure 4.10 – Arrimage des mauvais candidats .....	119
Figure 4.11 – Intercalation des meilleurs candidats.....	120
Figure 4.12 – Intercalation des mauvais candidats .....	121

## Liste des tableaux

Tableau I – Énergies d'excitation expérimentales ( $\lambda_{\text{exp}}$ ) et calculées ( $\lambda$ ) et forces d'oscillateur (F) des molécules <b>2.1</b> à <b>2.4</b> .....	43
Tableau II – Sommaire des différences d'énergies entre l'état de transition et l'état fondamental ( $\Delta E_{\text{Tr}}$ ) et de la comparaison de ces différences ( $\Delta\Delta E_{\text{Tr}}$ ). .....	47
Tableau III – Variation d'hybridation des atomes centraux aux angles $\alpha$ , $\beta$ , $\gamma$ et $\delta$ pour l'effet de charge, l'effet de la nature des atomes donneurs et accepteurs et l'effet de substituants. ....	52
Tableau IV – Facteurs de compressibilité utilisés au cours des dynamiques .....	75
Tableau V – Spécifications des systèmes du trimère .....	79
Tableau VI – Spécification des systèmes biphasiques .....	90
Tableau VII – Comparaison entre les valeurs $\gamma$ et les distances entre les groupements fonctionnels. ....	122



## Liste des abréviations

AM1	Austin Model 1
AMBER	Assisted Model Building with Energy Refinement
B3LYP	Becke 3-parameters, Lee-Yang-Parr
CGENFF	CHARMM General Force Field
CHARMM	Chemistry at HARvard Molecular Mechanics
CNDO	Complete Neglect of Differential Overlap
DFT	Density Functional Theory
FITTED	Flexibility Induced Through Targeted Evolutionary Description
GAFF	Generalized AMBER Force Field
GGA	Generalized Gradient Approximation
GROMACS	GRONingen Machine for Chemical Simulations
HF	Hartree-Fock
HOMO	Highest Occupied Molecular Orbital
INDO	Intermediate Neglect of Differential Overlap
ITO	Indium tin oxide
LDA	Local Density Approximation
LUMO	Lowest Unoccupied Molecular Orbital
MC-SCF	Multi-Configurational Self-Consistent field
MMFF94	Merck Molecular Force Field
MNDO	Modified Neglect of Differential Overlap
MOPAC2009	Molecular Orbital PACKage
MP2	Théorie de la perturbation de Moller-Plesset au deuxième ordre
NPT	Ensemble à particules (N), pression (P) et température (T) constante
NVT	Ensemble à particules (N), volume (V) et température (T) constante
OPLS	Optimized Potentials for Liquid Simulations
PBC	Periodic Boundary Conditions
PBE0	Perdew-Burke-Ernzerhoff

PBS	Phosphate Buffered Saline
PM3, PM5	Parameterized Model number 3, Parameterized Model number 5
PME	Particle Mesh Ewald
RCSB	Research Collaboratory for Structural Bioinformatics
RM1	Recife Model 1
STQN	Synchronous Transit-Guided Quasi-Newton
TD-DFT	Time-Dependant Density Functional Theory
THF	Tetrahydrofurane
UA	Unités arbitraires
VDW	van der Waals
ZDO	Zero Differential Overlap

*À ma famille.*

## Remerciements

Je tiens avant tout à remercier mes parents Jean-Pierre et Louise, mes frères Pierre-Olivier et Jordane et ma conjointe Kim pour m’avoir encouragé à poursuivre mon parcours universitaire et pour m’avoir soutenu dans mes victoires et mes échecs. Par votre écoute et votre dévouement, vous avez su me faire garder le cap.

Ce mémoire n’ayant pu être possible sans le travail rigoureux des expérimentalistes qui ont croisé ma route, je tiens à souligner la contribution de Pierre-Louis Brunner, d’Alice Heskia et de toute l’équipe du groupe Wuest pour leur partage de données expérimentales et des concepts de ces projets et pour leur aide dans les analyses par DFT. De la même manière, je souhaite souligner la contribution de Satu Strandman, de Christine Chhun, du professeur Julian Zhu et des autres membres des groupes Zhu et Schmitzer pour leur contribution aux projets de dynamique moléculaire. Finalement, bien qu’il ne me soit possible de les nommer, je ne peux passer sous silence la collaboration de nos collègues du domaine privé en ce qui concerne les projets touchant l’arrimage moléculaire. Je tiens par ailleurs à souligner la qualité du support technique de l’équipe du professeur Nicolas Moitessier développant FITTED et plus particulièrement de l’aide fournie par Éric Therrien.

J’aimerais remercier les professeurs Alain Rochefort, Matthias Ernzerhof et mes directeurs de recherches Andreea Schmitzer et James Wuest pour m’avoir aidé à renforcer mes connaissances en modélisation moléculaire et en chimie computationnelle en générale.

Finalement, aux membres du jury, je vous dis merci pour le temps consacré à l’évaluation et à la critique de cet ouvrage. J’espère que vous trouverez autant de plaisir à lire ce mémoire que j’en ai eu à l’écrire.

# Chapitre 1: Introduction

## 1.1. Les mathématiques et la connaissance

C'est en 1623, dans *Il Saggiatore*, que l'astronome et physicien Galileo Galilei écrivit la célèbre phrase « le livre de la Nature est écrit dans le langage des mathématiques »<sup>1</sup>. Plus de 300 ans plus tard, cet axiome est aujourd'hui endossé par les plus grands physiciens. Le mathématicien et philosophe Bertrand Russell dira des mathématiques qu'elles « ne possèdent pas seulement la Vérité, mais aussi la beauté suprême ». Albert Einstein, dans un discours prononcé à l'Académie des Sciences de Berlin en 1921, considérera que contrairement à toutes les propositions scientifiques qui peuvent être renversées et contredites par les expériences, les propositions mathématiques jouissent du fait qu'elles sont « absolument certaines et indisputables ».

Les modèles mathématiques, utilisés dans toutes les sciences, sont quant à eux définis comme « une représentation des aspects essentiels d'un système existant ou en devenir et qui présentent la connaissance sous une forme utilisable »<sup>2</sup>. Ils sont ainsi la représentation rationnelle de ce que nous considérons comme étant la réalité.

De nos jours, toute forme de progrès scientifique passe systématiquement par la mise en place de modèles mathématiques rigoureux permettant de rationaliser les observations. On voit notamment leur implication dans de multiples domaines purs et appliqués. Par exemple, la connaissance des équations décrites par la mécanique newtonienne permet aux ingénieurs de concevoir virtuellement de nouveaux objets et de simuler leurs propriétés (Figure 1.1) avant même de les construire<sup>3</sup>.

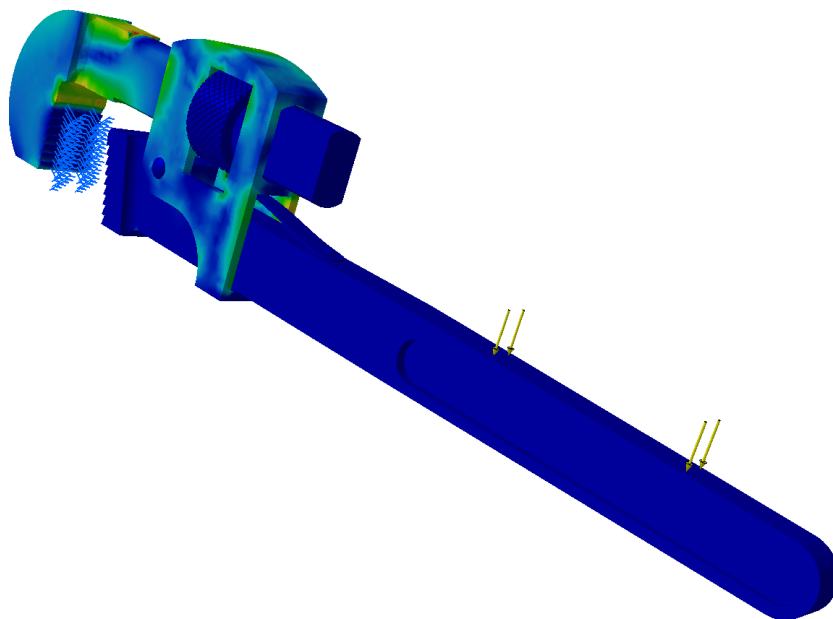


Figure 1.1 – Étude par éléments finis d’une pièce mécanique. Les flèches indiquent les forces appliquées sur l’objet. Image gracieusement fournie par son concepteur\*.

Les questions qui seront posées dans ce mémoire concernent principalement l’applicabilité des différents modèles mathématiques en chimie. Nous verrons, à travers différents cas réels, comment des systèmes simples peuvent être simulés par des méthodes computationnelles rigoureuses et quelles approximations peuvent être réalisées afin de simuler de très gros systèmes moléculaires et des situations complexes. La conception logique de cette étude présentera d’abord les modèles mathématiques les plus rigoureux et fera, d’un chapitre à l’autre, place aux approximations et aux modèles simplifiés pouvant représenter des systèmes moléculaires de plus en plus complexes et étendus.

Cette introduction vise à initier le lecteur aux différents modèles mathématiques utilisés en chimie computationnelle et en modélisation moléculaire. Sans avoir la prétention de nous

---

\* Pierre-Olivier Dubois, École Polytechnique de Montréal, Montréal, Québec H3C 3A7, Canada

adresser à des théoriciens, nous verrons dans ce chapitre le formalisme mathématique nécessaire à la représentation de la réalité à diverses échelles, et les approximations que nous pouvons faire pour simplifier ces modèles.

## **1.2. L'évolution des modèles mathématiques en chimie**

À mesure que les connaissances scientifiques générales se sont étoffées, les modèles atomiques proposés se sont raffinés. Nous noterons cependant quelques événements clés qui ont amorcé l'ère de la mécanique quantique actuelle.

### **1.2.1. La découverte de l'électron et le modèle *Plum Pudding***

En 1897, Sir Joseph John Thomson pu démontrer que les « rayons cathodiques » observés lors de l'application d'une différence de potentiel aux bornes de deux électrodes séparées par du vide étaient constitués de particules subatomiques négativement chargées. Cette première découverte jointe avec la découverte de l'existence des rayons X (1895) et de la radioactivité (1896) mettra en lumière la complexité des entités atomiques. Ces observations seront d'ailleurs les premiers obstacles sérieux à la physique classique qui prévalait à l'époque<sup>4</sup>. Cette découverte conduira notamment à l'élaboration du modèle atomique *Plum Pudding* en 1904, où un atome est composé d'électrons entourés d'une charge positive.

### **1.2.2. Le modèle Rutherford-Bohr**

En 1911, à la demande de Thomson, Ernest Rutherford se penche sur le modèle *Plum Pudding* afin de le valider. Alors qu'il bombarde de particules  $\alpha$  une mince feuille d'or, il constate qu'une particule sur 20 000 est déviée avec un angle d'environ  $90^\circ$ <sup>5</sup>, les autres passant directement à travers la feuille. Considérant que le modèle de l'atome plein ne

pouvait expliquer que la majorité des particules puissent passer à travers la feuille d'or, il suggère un modèle atomique planétaire, où les atomes seraient plutôt composés d'un noyau positif et dense avec des électrons gravitant autour, le tout séparé par du vide.

Ce nouveau modèle ne permettait toutefois pas d'expliquer la stabilité des orbitales électroniques. En effet, la loi de force de Coulomb est définie comme :

$$\frac{e^2}{4\pi\epsilon_0 r^2} = \frac{mv^2}{r} \quad (1.1)$$

où  $\epsilon_0$  est la permittivité du vide,  $m$  la masse,  $v$  la vitesse et  $r$  le rayon de l'orbite de l'électron. En supposant que l'électron tourne autour du noyau, la physique classique ne permettait pas d'expliquer comment un électron qui émet une radiation électromagnétique et perd de l'énergie ne va pas s'écraser sur le noyau, empêchant par conséquent l'existence d'une orbite électronique stable.

Le physicien Niels Bohr améliorera ce modèle en 1913 en supposant que les électrons évoluaient dans des orbitales précises, et que le moment angulaire des électrons était quantifié selon l'équation suivante :

$$mvr = n\hbar \quad n = 1, 2, 3, \dots \quad (1.2)$$

où  $\hbar = h/2\pi$ . En substituant dans l'équation 1.1, on obtient :



$$r = \frac{4\pi\epsilon_0\hbar^2 n^2}{me^2} \quad n = 1, 2, 3, \dots \quad (1.3)$$

impliquant donc des rayons permis discrets. Ce nouveau modèle put d'ailleurs raisonnablement expliquer l'existence des raies du spectre atomique de l'hydrogène.

### 1.2.3. Le comportement ondulatoire de la matière

En 1924, face au concept de la dualité onde-particule de la lumière, le scientifique Louis de Broglie eut l'idée de considérer la matière non pas seulement comme une particule, mais comme une entité pouvant aussi posséder un comportement ondulatoire<sup>4</sup>. Il utilisa les travaux d'Einstein sur la théorie de la relativité pour considérer que la lumière et la matière obéissaient à l'équation :

$$\lambda = \frac{h}{mv} \quad (1.4)$$

où une particule de masse  $m$  avec une vitesse  $v$  possédera une longueur d'onde de de Broglie de  $\lambda$ .

On pourra voir que pour les objets de grande masse (objets macroscopiques), cette longueur d'onde sera petite au point d'être indétectable et n'aura pour ainsi dire aucun impact sur le système. Par contre, à mesure que le système diminuera de taille (objet microscopique), une longueur d'onde résiduelle sera détectable.

### 1.3. L'équation de Schrödinger

Les contributions et avancées importantes de la physique pendant la première partie du XX<sup>e</sup> siècle avaient permis de rompre de manière définitive avec la physique classique. En 1926, le physicien Erwin Schrödinger publiera<sup>6</sup> sa célèbre équation visant à unifier le comportement des particules atomiques et subatomiques. L'équation de Schrödinger est un des piliers de la mécanique quantique. La forme décrivant les états stationnaires est définie par :

$$\hat{H}\psi(\vec{r}) = E\psi(\vec{r}) \quad (1.5)$$

où  $E$  est l'énergie de l'état stationnaire  $\psi(\vec{r})$ . Pour un système à  $N$  particules, l'Hamiltonien  $\hat{H}$  est constitué d'une partie cinétique  $\hat{T}$  et potentielle  $\hat{V}$  pour chaque  $i$  particule :

$$\hat{H} = \underbrace{\frac{-\hbar^2}{2} \sum_{i=1}^N \frac{\Delta_i}{m_i}}_{\hat{T}} + \underbrace{\frac{1}{4\pi\epsilon_0} \sum_{i=1, j>i}^N \frac{e^2 Z_i Z_j}{|\vec{r}_i - \vec{r}_j|}}_{\hat{V}} \quad (1.6)$$

Où  $\Delta_i$  est le Laplacien (un opérateur différentiel) appliqué à la position  $i$ ,  $\epsilon_0$  est la constante de permittivité du vide,  $Z_i$  la charge de la particule  $i$  et  $|\vec{r}_i - \vec{r}_j|$ , la distance entre les particules  $i$  et  $j$ . On peut voir que l'interaction entre les particules est définie par le deuxième terme puisque ce n'est que ce terme qui fait intervenir deux particules ( $i$  et  $j$ ). Pour un système à quatre particules (par exemple la molécule d'hydrogène), on devrait s'attendre à avoir un Hamiltonien à 10 termes (4 termes cinétiques + 6 termes de potentiels).

#### 1.4. L'approximation de Born-Oppenheimer

Il semble évident que pour des systèmes de taille conséquente, l'Hamiltonien devienne rapidement complexe et, par conséquent, que la fonction d'onde devienne pratiquement impossible à trouver. Une première simplification nommée « approximation de Born-Oppenheimer » fut suggérée en 1927 par les physiciens J. Robert Oppenheimer et Max Born<sup>7</sup>. Celle-ci se base sur le fait que les protons et les électrons possèdent des charges électriques de mêmes grandeurs, mais des masses très différentes (un proton est environ 1800 fois plus massif qu'un électron). Il en résulte donc que les vitesses des protons seront beaucoup plus faibles. L'approximation de Born-Oppenheimer néglige donc ces mouvements et considère les noyaux atomiques comme étant fixes. Cette considération aura des conséquences importantes sur la philosophie derrière la chimie computationnelle. De ce fait, la fonction d'onde ne sera plus réellement une fonction d'onde du système, mais plutôt une fonction d'onde électronique. Ainsi, deux avenues seront considérées lorsqu'un calcul sera réalisé :

1. La mécanique quantique : La mécanique quantique telle que nous l'utiliserons dans le prochain chapitre est, à proprement parler, l'étude des fonctions d'onde et des densités électroniques. Elle utilise diverses théories qui seront présentées dans les sections suivantes pour déterminer comment les électrons des molécules s'organisent et interagissent.
2. La mécanique moléculaire : La mécanique moléculaire telle que nous l'utiliserons dans les chapitres 3 et 4 est l'étude des interactions des noyaux atomiques. Elle utilise des modèles, la plupart du temps empiriques, pour simuler les mouvements des noyaux atomiques et ne considère pas *a priori* les électrons comme éléments

quantiques. Nous verrons que la mécanique moléculaire ne répond pas aux questions quantiques, mais rend plutôt le caractère classique aux molécules.

## 1.5. Les calculs de structure électronique

### 1.5.1. La théorie Hartree-Fock

Le développement de l'Hamiltonien pour des systèmes multiélectroniques conduit à des complexités mathématiques qui empêchent l'obtention d'une fonction d'onde de manière analytique. Pour contourner ce problème, Douglas Hartree, Vladimir Fock et John C. Slater ont approximé la fonction d'onde pour  $N$  électrons par un déterminant de Slater :

$$\psi(\vec{r}_1, \dots, \vec{r}_N) \cong \frac{1}{\sqrt{N!}} \begin{vmatrix} \varphi_1(\vec{r}_1) & \dots & \varphi_N(\vec{r}_1) \\ \vdots & \ddots & \vdots \\ \varphi_1(\vec{r}_N) & \dots & \varphi_N(\vec{r}_N) \end{vmatrix} \quad (1.7)$$

où  $\varphi_N$  représente une spin-orbitale. L'application de l'opérateur  $\hat{H}$  donne l'énergie Hartree-Fock selon :

$$E^{HF} = \sum_{i=1}^N \langle \varphi_i | h | \varphi_i \rangle + \underbrace{\sum_{i,j} \langle \varphi_i \varphi_j | V | \varphi_i \varphi_j \rangle}_{\text{terme direct}} - \underbrace{\sum_{i,j} \langle \varphi_i \varphi_j | V | \varphi_j \varphi_i \rangle}_{\text{terme d'échange}} \quad (1.8)$$

Dans la sommation précédente, trois termes sont indiqués. Le premier réfère à l'énergie cinétique et potentielle des électrons due à l'interaction avec les noyaux et consiste en un terme monoélectronique puisqu'il ne se réfère qu'à un seul électron (il n'y a que l'indice  $i$ ).

Les deux autres termes sont multiélectroniques puisque les sommations réfèrent aux indices  $i$  et  $j$ . Le terme direct représente l'énergie de la répulsion électrostatique de la densité électronique avec elle-même. Le terme d'échange ne possède pas de réelle signification physique, mais consiste en une correction de l'énergie due à l'antisymétrie de la fonction d'onde.

### **1.5.2. Les méthodes Hartree-Fock semi-empiriques**

Comme nous l'avons vu, le formalisme de Hartree-Fock permet de calculer l'énergie des molécules en fonction des orbitales. Ces informations permettent de déduire et de comprendre certains comportements moléculaires découlant des interactions atomiques.

Malgré les approximations déjà faites, l'évaluation de certaines intégrales demeure tout de même extrêmement pénible et longue à calculer, en particulier pour des molécules de grande taille. Cette difficulté est en partie attribuable à la présence d'une double sommation pour le calcul des termes directs et des termes d'échange. Ainsi, on peut voir que l'ordre de l'équation 1.8 évolue en  $O(N^4)$  pour  $N$  fonctions de base représentant une combinaison linéaire de la fonction d'essai  $\varphi^8$ .

Pour améliorer la vitesse des calculs, plusieurs simplifications sont effectuées dans les méthodes Hartree-Fock semi-empiriques. En premier lieu, seuls les électrons de valence sont considérés. Ceux-ci sont modélisés comme évoluant dans des champs initiés par des particules fixes composées des noyaux et des électrons de cœur. Conséquemment, les interactions noyaux-noyaux sont évaluées de manière différente. Pour la méthode MNDO<sup>9</sup>, l'interaction est calculée avec l'équation :

$$V_{nuc-nuc}(i,j) = Z_i Z_j \langle ii|jj \rangle (1 + e^{-\alpha_i R_{ij}} + e^{-\alpha_j R_{ij}}) + \frac{Z_i Z_j}{R_{ij}} \left( \sum_k a_{ki} e^{-b_{ki}(R_{ij}-c_{ki})^2} + \sum_k a_{kj} e^{-b_{kj}(R_{ij}-c_{kj})^2} \right) \quad (1.9)$$

où  $Z_i$  représente le nombre d'électrons de valence du noyau  $i$ ,  $R_{ij}$  la distance entre les noyaux  $i$  et  $j$ , et  $\alpha_i$ ,  $a_{ki}$ ,  $b_{ki}$ ,  $c_{ki}$  des paramètres ajustables. Le paramètre  $k$  varie entre 2 et 4 en fonction du type d'atome. Le deuxième terme, présent seulement dans la méthode AM1<sup>10</sup>, a été rajouté pour corriger les répulsions excessives de noyaux à grande distance.

En deuxième lieu, des bases minimales sont considérées pour décrire les orbitales moléculaires. Dans la plupart des méthodes, le nombre minimal d'orbitales atomiques requis pour accommoder les électrons de l'atome neutre est considéré. La plupart des méthodes ne considèrent que les orbitales s et p<sub>x</sub>, p<sub>y</sub> et p<sub>z</sub>.

Finalement, certaines intégrales sont négligées. L'approximation *Zero Differential Overlap* (ZDO), utilisée dans les méthodes semi-empiriques, consiste à négliger certaines interactions entre des électrons des orbitales éloignées ou la contribution électronique dans des orbitales atomiques de types différents.

Par exemple, considérons l'intégrale biélectronique suivante à solutionner :

$$\langle \mu\nu | \lambda\sigma \rangle = \iint \chi_\mu^A(1) \chi_\nu^B(2) V \chi_\lambda^C(1) \chi_\sigma^D(2) d\vec{r}_1 d\vec{r}_2 \quad (1.10)$$

où  $\chi_\mu^A(1)$  représente la fonction pour l'électron 1 de type  $\mu$  (ex : orbitale s ou p) positionné sur l'atome A. L'approximation ZDO stipule qu'il est possible de négliger le produit des fonctions  $\chi_\mu^A(1)\chi_\nu^B(1)$  lorsque  $\mu \neq \nu$ . Cette approximation est utilisée de différente manière en fonction des méthodes semi-empiriques.

Les méthodes de type CNDO<sup>11</sup> utilisent l'approximation de manière systématique pour toutes les intégrales. Les méthodes de type INDO<sup>12</sup> négligent toutes les intégrales de recouvrement à l'exception du cas où  $A = B = C = D$  dans l'équation 1.10, ce qui correspond au cas où les quatre fonctions se trouvent sur le même atome. Les méthodes de type MNDO<sup>9</sup> ne négligent pas les intégrales où  $A = B$  et  $C = D$ , c'est à dire lorsque les fonctions de base du premier électron sont situées sur le même atome et que les fonctions de base du second électron sont situées sur le même atome. Cette approximation permet de passer d'un problème d'ordre  $O(N^4)$  à un problème d'ordre  $O(N^2)$ <sup>8</sup>.

### 1.5.3. La théorie de la fonctionnelle de densité

Le formalisme de Hartree-Fock constitue une méthode intéressante pour étudier des systèmes moléculaires. Toutefois, la complexité mathématique nous oblige à faire un compromis entre les approximations et l'exactitude des résultats. Pour contourner ce problème, Pierre Hohenberg, Walter Kohn et Lu Sham ont développé, dans les années soixante, la théorie de la fonctionnelle de densité visant à remplacer les fonctions d'ondes multiélectroniques, qui dépendent de  $3N$  variables spatiales (pour  $N$  particules), par une fonctionnelle de la densité, qui ne dépend que de trois variables spatiales. Le principal défi de ce nouveau modèle reste que cette fonctionnelle est inconnue. On définit la densité  $\rho(\vec{r})$  comme la probabilité de trouver n'importe quel des  $N$  électrons avec un spin arbitraire dans un volume en présence des autres électrons<sup>13</sup>.

#### 1.5.4. Le modèle Kohn-Sham

Pour tenter de modéliser cette fonctionnelle, Kohn et Sham eurent l'idée de décomposer l'énergie en plusieurs fonctionnelles distinctes<sup>14</sup>. Les théorèmes de Hohenberg-Kohn postulent que la minimisation d'une fonctionnelle permet d'obtenir l'énergie de l'état fondamental d'un système<sup>15</sup>:

$$E_0 = \min \left( F[\rho] + \int \rho(\vec{r}) V_{Ne} d\vec{r} \right) \quad (1.11)$$

où  $F[\rho]$  est une fonctionnelle contenant l'énergie cinétique, les interactions coulombiennes et les autres termes non classiques tels que l'échange et les effets de corrélation et le second terme, l'énergie d'interaction entre les noyaux et les électrons. La fonctionnelle  $F[\rho]$  est donc définie telle que :

$$F[\rho(\vec{r})] = T_s[\rho(\vec{r})] + J[\rho(\vec{r})] + E_{xc}[\rho(\vec{r})] \quad (1.12)$$

Les interactions coulombiennes  $J[\rho(\vec{r})]$  peuvent être calculées de manière exacte. L'énergie cinétique  $T_s[\rho(\vec{r})]$  n'est pas connue, mais est calculée par les orbitales de Kohn-Sham. Finalement, le dernier terme  $E_{xc}[\rho(\vec{r})]$  est inconnu et doit être approximé. On obtient au final la forme suivante :



$$\begin{aligned}
F[\rho(\vec{r})] = & \underbrace{-\frac{1}{2} \sum_i^N \langle \varphi_i | \nabla^2 | \varphi_i \rangle}_{T_s[\rho]} \\
& + \underbrace{\frac{1}{2} \sum_i^N \sum_j^N \iint |\varphi_i(\vec{r}_1)|^2 \frac{1}{|\vec{r}_1 - \vec{r}_2|} |\varphi_j(\vec{r}_2)|^2 d\vec{r}_1 d\vec{r}_2}_{J[\rho]} \\
& - \underbrace{\sum_i^N \int \sum_A^M \frac{Z_A}{|\vec{r}_1 - \vec{r}_A|} |\varphi_i(\vec{r}_1)|^2 d\vec{r}_1}_{\int \rho(\vec{r}) V_{Ne} d\vec{r}} + E_{xc}[\rho(\vec{r})]
\end{aligned} \tag{1.13}$$

L'approximation locale de la densité, l'approximation des gradients généralisée et les méthodes hybrides permettent de d'approximer l'énergie d'échange et de corrélation.

### 1.5.5. Approximation locale de la densité

L'approximation locale de la densité (LDA) se base sur le modèle du gaz uniforme d'électrons. Il s'agit d'un système où des électrons évoluent dans un milieu chargé positivement. Ce système est physiquement comparable à une structure cristalline parfaite, mais s'éloigne grandement de la réalité d'un système moléculaire plus complexe où les densités peuvent grandement varier sur l'ensemble d'une molécule. L'énergie d'échange-corrélation de la LDA est définie comme :

$$E_{xc}^{LDA}[\rho] = \int \rho(\vec{r}) \left( -\frac{3}{4} \sqrt{\frac{3\rho(\vec{r})}{\pi}} + E_c(\rho(\vec{r})) \right) d\vec{r} \tag{1.14}$$

Il n'existe pas d'expression explicite pour l'énergie de corrélation  $E_c(\rho(\vec{r}))$ ; toutefois, des simulations numériques permettent d'obtenir de manière précise ces valeurs<sup>16</sup>.

### 1.5.6. Approximation des gradients généralisée

L'approximation des gradients généralisée (GGA) utilise non seulement l'information de la densité à un point particulier, mais aussi le gradient de celle-ci, de sorte à pouvoir tenir compte de l'hétérogénéité de la densité électronique des molécules. Cette nouvelle approximation est obtenue en considérant l'approximation locale de la densité comme le premier terme d'une expansion de Taylor :

$$E_{XC}^{GGA}[n_{\uparrow}, n_{\downarrow}] = \int f(n_{\uparrow}, n_{\downarrow}, \nabla n_{\uparrow}, \nabla n_{\downarrow}) d\vec{r} \quad (1.15)$$

où  $f(\dots)$  est une fonctionnelle des densités de spin  $n_{\uparrow}$  et  $n_{\downarrow}$  et  $\nabla$  est le gradient de ces densités.

### 1.5.7. Méthodes hybrides

Nous avons vu que le terme d'échange-corrélation en DFT est un terme pour lequel nous n'avons pas de méthode de résolution exacte. Par contre, si les méthodes Hartree-Fock ne permettent pas de calculer la corrélation électronique, elles donnent une valeur plutôt exacte pour l'échange. Les méthodes hybrides proposent ainsi d'utiliser le terme d'échange des méthodes Hartree-Fock avec les termes d'échange et de corrélation des méthodes DFT pour calculer l'énergie d'échange-corrélation. Il s'agit donc d'un mélange entre les deux théories.

## 1.6. La mécanique moléculaire

La mécanique quantique est un outil puissant utilisé pour mieux comprendre les petits systèmes moléculaires et pour répondre à certaines questions concernant la structure électronique de molécules. Toutefois, lorsque les questions posées concernent des objets de taille conséquente ou bien demandent l'étude d'informations statistiques et structurales, les modèles mathématiques présentés précédemment deviennent rapidement beaucoup trop lourds à traiter.

Comment peut-on dès lors simuler des systèmes contenant des milliers d'atomes interagissant entre eux ou des systèmes dynamiques évoluant sur des échelles temporelles conséquentes? Un retour aux modèles mathématiques classiques est de mise. En effet, en décrivant les molécules en fonction de leurs noyaux et non pas de leurs électrons, la mécanique moléculaire définit une fonction purement empirique prenant pour variable, les positions des noyaux et impliquant de manière implicite la présence des électrons.

La mécanique moléculaire considère une molécule comme une série de masses (les atomes) reliés entre eux par des ressorts (liaisons chimiques). En fonction des interactions entre les masses, des déformations structurales peuvent avoir lieu. Plus particulièrement, pour l'ensemble des méthodes, l'énergie du système est décrite par la forme fonctionnelle suivante :

$$E = \underbrace{E_{lien} + E_{angle} + E_{dièdre}}_{E_{covalente}} + \underbrace{E_{coul} + E_{LJ}}_{E_{non-covalente}} \quad (1.16)$$

où  $E_{lien}$  est l'énergie des liens covalents entre les atomes du système,  $E_{angle}$  est l'énergie des angles formés entre les atomes reliés, et  $E_{dièdre}$  est l'énergie des angles dièdres des

atomes. Toutes ces composantes énergétiques permettent de caractériser une interaction covalente particulière. Pour simuler correctement les interactions non covalentes entre les atomes, deux termes supplémentaires sont ajoutés.  $E_{coul}$  est l'énergie coulombienne entre les atomes non reliés et finalement  $E_{LJ}$  est l'énergie découlant d'un potentiel de Lennard-Jones. Les modèles de tous ces types d'interactions seront décrits dans les sous-sections suivantes.

### 1.6.1. Potentiel de liaison

Le potentiel de liaison est l'énergie découlant de l'étirement du lien covalent entre deux atomes. La forme fonctionnelle est définie par :

$$E_{Lien} = k_{ij}(l_{ij} - l_0)^2 \quad (1.17)$$

où  $k_{ij}$  est la constante de force du ressort,  $l_0$  la longueur d'équilibre du ressort et  $l_{ij}$  la distance entre les atomes  $i$  et  $j$ .

Le modèle mathématique de l'oscillateur harmonique permet de décrire avec précision le phénomène d'élongation d'un lien près de la distance d'équilibre, mais devient rapidement faux lorsque la distance entre les atomes devient grande et que le lien covalent se brise (Figure 1.2). La mécanique moléculaire ne permet donc pas de décrire des systèmes moléculaires où le bris ou la formation de lien chimique survient.

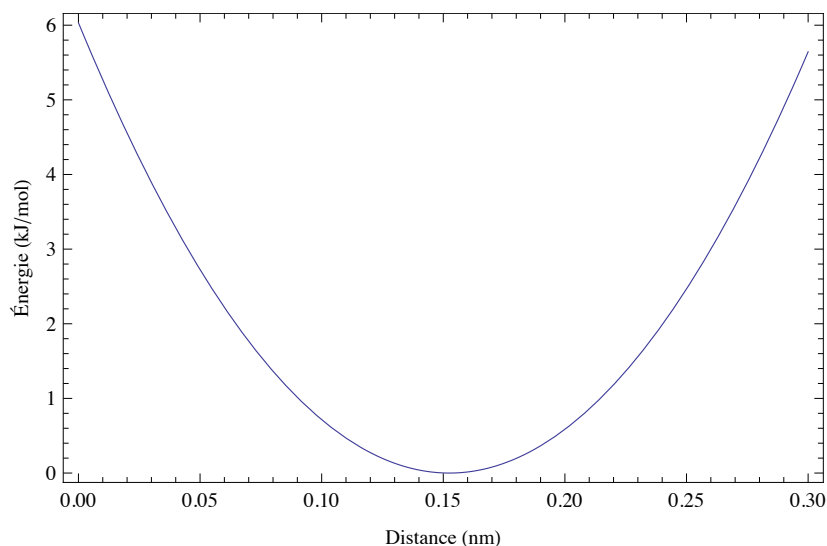


Figure 1.2 – Oscillateur harmonique décrivant l'énergie d'étirement d'un lien covalent. La distance d'équilibre dans cet exemple est de 0,15 nm.  $k_{ij}$  est égal à 259,408 kJ/(mol · nm<sup>2</sup>).

### 1.6.2. Potentiel d'angle

De la même manière, l'angle entre trois atomes  $i$ ,  $j$  et  $k$  avec l'atome central  $j$  est défini par l'oscillateur harmonique. La forme fonctionnelle est la suivante :

$$E_{Angle} = k_{ijk}(\theta_{ij-jk} - \theta_0)^2 \quad (1.18)$$

où  $k_{ijk}$  est la constante de déformation de l'angle entre les liens  $ij$  et  $jk$ ,  $\theta_0$  est l'angle d'équilibre des liens et  $\theta_{ij-jk}$  est l'angle entre les atomes  $i$ ,  $j$  et  $k$ .

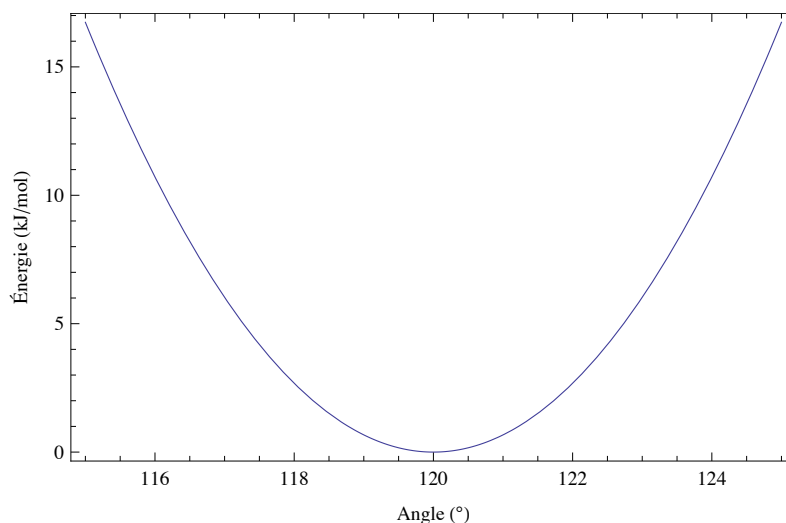


Figure 1.3 – Oscillateur harmonique décrivant l'énergie de déformation de l'angle entre deux liens covalents. L'angle d'équilibre dans cet exemple est de 120°.  $k_{ijk}$  est égal à 669,440 kJ/mol · (°)<sup>2</sup>.

### 1.6.3. Potentiel d'angle dièdre

L'angle dièdre est formé par les plans passant par les atomes  $i, j, k$  et  $j, k, l$ . Il permet d'exprimer plusieurs comportements moléculaires tels que les rapports de conformation *cis* et *trans* et les tensions de cycle. La forme fonctionnelle est définie par l'expansion en série de cosinus de Fourier :

$$E_{dièdre} = \frac{1}{2} \sum_n^N V_n (1 + \cos(n\varphi_{ijk-jkl} - \gamma_n)) \quad (1.19)$$

où  $V_n$  et  $n$  sont des constantes d'angles,  $\gamma_n$  est la phase du terme  $n$  et  $\varphi_{ijk-jkl}$  est l'angle dièdre. L'utilisation d'une série permet d'exprimer à peu près n'importe quel type de potentiel périodique sur un domaine variant entre 0 et  $2\pi$ <sup>17</sup>, ce qui est pratique pour définir des interactions aussi complexes et variées.

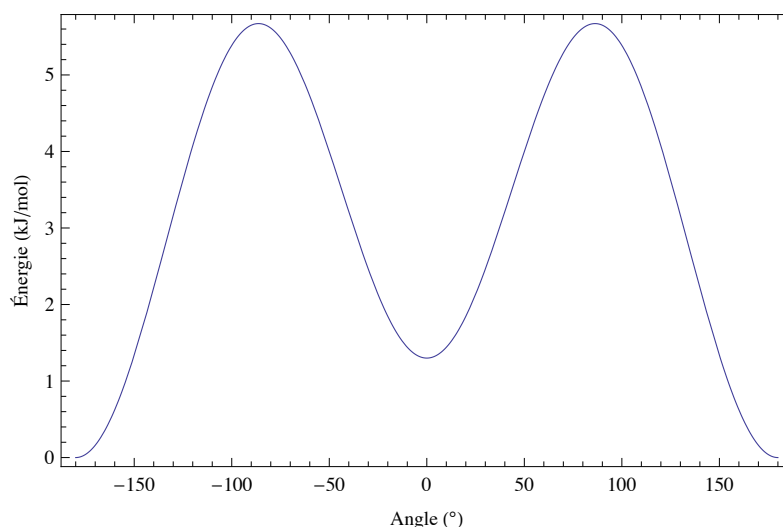


Figure 1.4 – Fluctuation de l'énergie en fonction de l'angle dièdre H-N-C-O d'une liaison peptidique. Il est possible de voir dans cet exemple la stabilisation de la conformation *trans* (180°) par rapport à la conformation *cis* (0°).

#### 1.6.4. Potentiel électrostatique

La forme fonctionnelle du potentiel électrostatique est définie par :

$$E_{Coul} = \frac{q_i q_j}{4\pi\epsilon r_{ij}} \quad (1.20)$$

où  $q_i$  et  $q_j$  sont les charges partielles des atomes  $i$  et  $j$ ,  $\epsilon$  la permittivité du milieu et  $r_{ij}$  la distance entre les atomes. Les charges partielles n'ont pas de sens physique sinon celui de disparité entre la contribution des protons et des électrons. Cette disparité peut être calculée par les équations de la mécanique quantique. On peut voir comment les charges complémentaires vont naturellement abaisser l'énergie du système et comment les charges de même signe vont augmenter l'énergie.

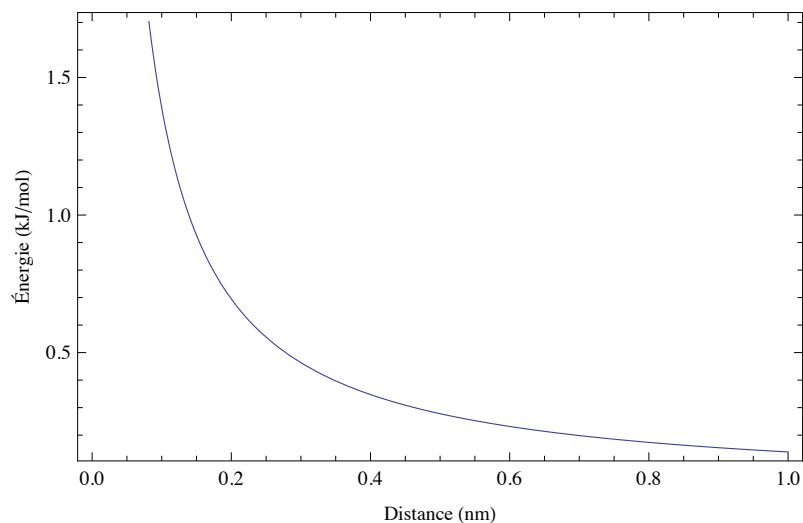


Figure 1.5 – Effet de la distance sur le potentiel électrostatique entre deux atomes partiellement chargés positivement.

### 1.6.5. Potentiel de Lennard-Jones

Finalement, la mécanique moléculaire inclut un dernier terme pour considérer les interactions entre une paire d'atomes neutres. Le modèle utilisé pour simuler ce type d'interaction est celui du potentiel de Lennard-Jones<sup>18</sup>. Décrite en 1924, la forme fonctionnelle du potentiel de Lennard-Jones est la suivante :

$$E_{LJ} = 4\varepsilon \left( \left( \frac{\sigma}{r_{ij}} \right)^{12} - \left( \frac{\sigma}{r_{ij}} \right)^6 \right) \quad (1.21)$$

où  $\varepsilon$  représente la profondeur du puits de potentiel et  $\sigma$  la distance des atomes à l'équilibre divisé par  $\sqrt[6]{2}$ . Graphiquement, le potentiel de Lennard-Jones prend la forme suivante :



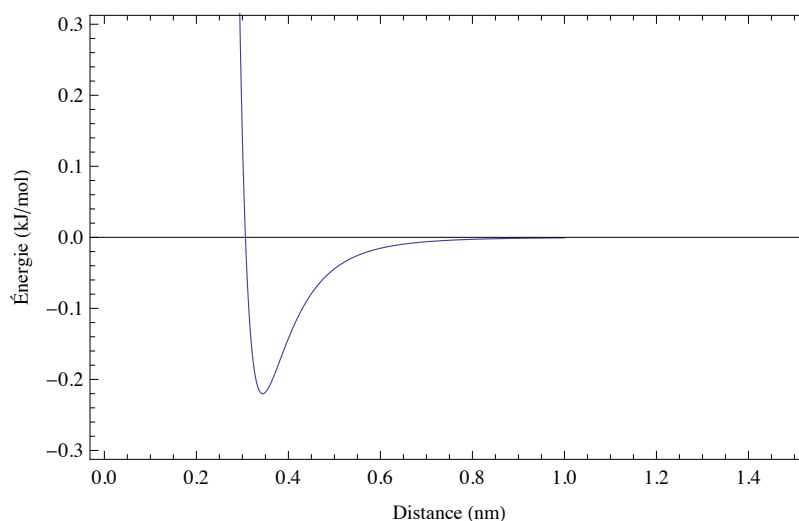


Figure 1.6 – Potentiel de Lennard-Jones entre deux atomes neutres.  $\sigma$  et  $\varepsilon$  sont égal à 0,307 nm et 0,880 kJ/mol respectivement.

Le premier terme (puissance 12) est un terme répulsif. Il vise à décrire l'intégrale d'échange de particules de même spin d'orbitales atomiques se chevauchant. Le second terme (puissance 6) est attractif et il vise à décrire les interactions de dispersion.

#### 1.6.6. Les champs de force

Comme nous l'avons vu, la mécanique moléculaire possède sa propre forme fonctionnelle découlant de modèles mathématiques empiriques pour décrire les systèmes moléculaires. Parmi la liste des variables géométriques requises pour calculer l'énergie conformationnelle des molécules, beaucoup sont propres à la nature des atomes et proviennent de paramètres découlant des études comparatives avec des expérimentations et des calculs quantiques. Ce regroupement de paramètres se nomme un champ de force, et est requis pour n'importe quel calcul impliquant la mécanique moléculaire. Une description plus étoffée des champs de force et de leurs forces et faiblesses respectives sera faite au Chapitre 3.

## Chapitre 2: Utilisation de la théorie de la fonctionnelle de la densité dans l'étude des interactions intermoléculaires

Les multiples modèles mathématiques brièvement décrits au chapitre précédent nous permettront, tout au long de cette étude, de prédire, comprendre et confirmer certains phénomènes moléculaires.

Dans ce chapitre, nous discuterons de l'applicabilité de la théorie Hartree-Fock et de la théorie de la fonctionnelle de la densité en chimie, principalement sur des systèmes moléculaires de petite taille. Ces méthodes, par leur approche quantique, apportent non seulement des informations sur le caractère structural des arrangements moléculaires, mais aussi sur le caractère électronique de la matière. Nous nous pencherons sur deux systèmes moléculaires. Premièrement, nous modéliserons les orbitales HOMO et LUMO de molécules photoactives par TD-DFT. Ensuite, nous observerons des interactions intramoléculaires par l'étude des états de transition de mésomères analogues au 1,3,5-triazapentadiène par DFT.

Tel que définie en introduction et de manière générale, la mécanique quantique est décrite selon deux schémas : la théorie Hartree-Fock et la théorie de la fonctionnelle de la densité. La théorie Hartree-Fock consiste à approximer la fonction d'onde antisymétrique par un déterminant de Slater (section 1.5.1). Cette approximation revient à considérer la probabilité simultanée  $P_{12}$  pour deux électrons aux positions  $\vec{r}_1$  et  $\vec{r}_2$  comme étant le produit de deux probabilités simples, chaque électron évoluant dans son orbitale (équation 2.1).

$$d^2P_{12}(\vec{r}_1, \vec{r}_2) = |\phi_1(\vec{r}_1)|^2 dv_1 |\phi_2(\vec{r}_2)|^2 dv_2 \quad (2.1)$$

La caractéristique principale de l'équation 2.1 est que les événements sont non corrélés, ce qui signifie que la probabilité d'un événement (ici de trouver l'électron 1 dans un volume  $dv_1$ ) n'est pas modifiée par la survenue de l'autre (la probabilité de trouver l'électron 2 dans un volume  $dv_2$ ). En pratique, pourtant, les électrons sont corrélés de deux façons : en tant que particules de même charge qui se repoussent (trou de Coulomb) et en tant que fermions qui se repoussent lorsqu'ils possèdent des spins identiques (trou de Fermi).

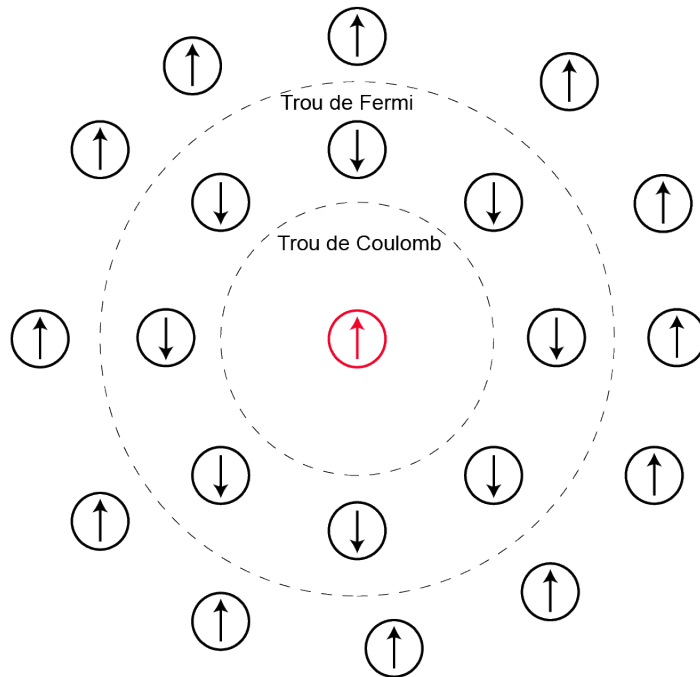


Figure 2.1 – Illustration d'un trou de Coulomb et d'un trou de Fermi autour d'un électron (en rouge).

La théorie Hartree-Fock tient compte du trou de Fermi en ce sens que l'antisymétrisation de la fonction d'onde empêche deux électrons de même spin de se situer au même point. Toutefois, le trou de Coulomb n'est pas pris en compte dans la méthode puisque deux

électrons de spin opposés peuvent cohabiter dans un même espace. On peut voir assez facilement ce comportement en calculant la probabilité que deux électrons se trouvent en un même point. En appliquant l'équation 1.7 pour deux électrons de même spin, on obtient :

$$\psi(\vec{x}_1, \vec{x}_2) = \frac{1}{\sqrt{2}} (\phi_1(\vec{r}_1)\alpha(\sigma_1)\phi_2(\vec{r}_2)\alpha(\sigma_2) - \phi_1(\vec{r}_2)\alpha(\sigma_2)\phi_2(\vec{r}_1)\alpha(\sigma_1)) \quad (2.2)$$

La densité de probabilité est obtenue en observant le carré de l'équation 2.2. En intégrant sur les coordonnées de spin, on obtient :

$$\begin{aligned} \psi(\vec{r}_1, \vec{r}_2)^2 = \frac{1}{\sqrt{2}} & ([\phi_1(\vec{r}_1)\phi_2(\vec{r}_2)]^2 + [\phi_1(\vec{r}_2)\phi_2(\vec{r}_1)]^2 \\ & - 2\phi_1(\vec{r}_1)\phi_2(\vec{r}_2)\phi_1(\vec{r}_2)\phi_2(\vec{r}_1)) \end{aligned} \quad (2.3)$$

Si on considère que les électrons sont au même endroit, on peut voir que  $\psi(\vec{r}_1, \vec{r}_1)^2 = 0$ . La méthode Hartree-Fock interdit à deux électrons de même spin de se situer au même endroit. Par contre, en réalisant la même démarche pour des électrons de spin différent, on obtient :

$$\psi(\vec{r}_1, \vec{r}_2)^2 = \frac{1}{\sqrt{2}} ([\phi_1(\vec{r}_1)\phi_2(\vec{r}_2)]^2 + [\phi_1(\vec{r}_2)\phi_2(\vec{r}_1)]^2) \quad (2.4)$$

Cette fois-ci, la densité de probabilité n'est pas nulle lorsque  $\vec{r}_2 = \vec{r}_1$ . Il en résulte donc que la probabilité de retrouver des électrons de spins opposés rapprochés est toujours surévaluée et que l'énergie du système est donc trop grande. La différence entre l'énergie de Hartree-Fock et l'énergie réelle est appelée énergie de corrélation<sup>19</sup>.

Une des solutions envisagées pour considérer le trou de Coulomb et traiter l'énergie de corrélation consiste à mélanger les états électroniques en approximant la fonction d'onde par plusieurs déterminants de Slater au lieu d'un seul. Certaines méthodes telles que l'interaction de configuration<sup>20</sup>, la MC-SCF<sup>21</sup> (Multi-Configurational Self-Consistent Field) et la théorie des perturbations de Møller-Plesset<sup>22</sup> permettent d'obtenir ce genre de fonctions d'onde, au prix toutefois d'un coût élevé en calcul.

## 2.1. Choix de la DFT

Tel que nous l'avons vu, les méthodes Hartree-Fock ne permettent pas d'obtenir l'énergie de corrélation du système, à moins d'utiliser de coûteuses méthodes computationnelles. La théorie de la fonctionnelle de la densité vient toutefois répondre à cette impasse en traitant de manière approximative l'énergie d'échange et de corrélation par l'usage d'une fonctionnelle approximative dédiée à cet effet. Le principal défi consistait donc à trouver une bonne fonctionnelle d'échange et de corrélation pour les systèmes qui nous intéressent. Étant donné la taille de nos systèmes, nous pouvions penser à utiliser les méthodes hybrides pour tous les calculs. À partir de ce choix, les deux fonctionnelles les plus populaires ont été considérées. Nous décrirons dans cette section les fonctionnelles B3LYP<sup>23</sup> et PBE0<sup>24</sup> et ce qui a motivé notre choix final.

### 2.1.1. B3LYP

La fonctionnelle hybride B3LYP est probablement l'une des plus utilisées en chimie computationnelle actuellement<sup>25</sup>. La fonctionnelle d'échange est définie telle que :

$$E_X^{B3LYP} = 0,8E_X^{LDA} + 0,2E_X^{HF} + 0,72\Delta E_X^{B88} \quad (2.5)$$

où  $E_X^{LDA}$  est l'énergie d'échange LDA,  $E_X^{LDA}$  est l'énergie d'échange Hartree-Fock et  $\Delta E_X^{B88}$  est la correction apportée par l'échange Becke88<sup>26</sup>. La fonctionnelle de corrélation est définie telle que :

$$E_C^{B3LYP} = 0,19E_C^{VWN3} + 0,81E_C^{LYP} \quad (2.6)$$

où  $E_C^{VWN3}$  et  $E_C^{LYP}$  font référence aux énergies de corrélation des fonctionnelles Vosko-Wilk-Nusair III<sup>27</sup> et Lee-Yang-Parr<sup>28</sup>. Les coefficients des deux dernières équations ont été définis de manière semi-empirique en comparant les énergies d'atomisation, les affinités électroniques et les potentiels d'ionisation expérimentaux selon la théorie Gaussian-2<sup>29</sup> pour des composés constitués d'atomes dans l'espace chimique entre l'hydrogène et le chlore.

La fonctionnelle B3LYP est reconnue pour donner, de manière globale, des géométries, énergies, déplacements chimiques et forces d'oscillateurs relativement fidèles aux observations expérimentales<sup>25, 30</sup>. Toutefois, étant donné que les coefficients des équations 2.5 et 2.6 ont été déterminés en paramétrant sur des composés d'un espace chimique restreint, les calculs sur des systèmes différents peuvent entraîner de très grandes variations

dans la qualité des résultats. Il a d'ailleurs été reconnu que B3LYP ne donnait pas de bons résultats pour les métaux<sup>30</sup>.

### 2.1.2. PBE0

La fonctionnelle hybride PBE0 est construite en mélangeant 25 % d'échange Hartree-Fock à 75 % de la fonctionnelle d'échange PBE<sup>31</sup>. La corrélation est obtenue en considérant la fonctionnelle de densité de PBE. Le terme d'énergie d'échange-corrélation de PBE0 s'écrit :

$$E_{XC}^{PBE0} = \frac{1}{4}E_X^{HF} + \frac{3}{4}E_X^{PBE} + E_c^{PBE} \quad (2.7)$$

Le ratio entre Hartree-Fock et DFT provient d'un modèle théorique<sup>32</sup> développé en utilisant la connexion adiabatique<sup>33</sup>. Celle-ci permet de connecter le système Kohn-Sham sans interaction au système réel.

Il s'agit donc d'une fonctionnelle dite non empirique puisque, contrairement à la fonctionnelle B3LYP (section 2.1.1), les paramètres n'ont pas été déterminés en comparant les modèles avec les données expérimentales ou avec d'autres fonctionnelles.

PBE0 a été utilisé pour tous les calculs de ce chapitre. Parmi les raisons motivant ce choix, nous retiendrons les suivantes :

- L'utilisation<sup>24a</sup> de la fonctionnelle PBE0 sur la même série de molécules que Gaussian-2 (section 2.1.1) montre des erreurs comparables à l'utilisation de la fonctionnelle B3LYP qui elle découle de plusieurs paramètres empiriques.
- Étant donné qu'il s'agit d'une fonctionnelle non-empirique, PBE0 permet de faire un bon compromis entre la qualité des calculs et le respect des principes physiques découlant du modèle présenté<sup>24</sup>.
- La fonctionnelle PBE0 est une théorie de choix dans l'étude des systèmes excités faisant intervenir la TD-DFT<sup>34</sup>. Il a été montré qu'elle permettait notamment de décrire les phénomènes physiques survenant lors d'une excitation photochimique et que les valeurs énergétiques des transitions HOMO-LUMO sont près des valeurs expérimentales<sup>34a</sup> et généralement meilleures ou similaires aux autres fonctionnelles<sup>34b</sup>. Nous aurons l'occasion de discuter plus amplement de la TD-DFT dans la sous-section 2.1.2.

### 2.1.1. Fonctions de base

En introduction, nous avons discuté de la théorie Hartree-Fock (section 1.5.1) et de la façon dont nous pouvions obtenir l'énergie du système à partir de fonctions d'essais. Nous n'avons toutefois pas défini celles-ci autrement que comme étant une combinaison linéaire de  $N$  fonctions de base  $\phi$ . Nous décrirons plus en détail la nature de ces fonctions  $\phi$  dans le cadre de cette sous-section.

Les fonctions de base sont généralement des orbitales atomiques centrées sur un atome. La combinaison linéaire de ces fonctions forme une fonction d'essai qui est par la suite utilisée dans le déterminant de Slater (Équation 1.7) :



$$\varphi = \sum_i c_i \phi_i \quad (2.8)$$

Dans la majorité des cas, les fonctions de base sont approximées par une combinaison linéaire de gaussiennes. Dans les descriptions les plus sommaires des systèmes, on se contentera d'utiliser le nombre minimal de gaussiennes nécessaire à la représentation de tous les électrons de chaque atome. Par exemple, un atome de carbone pourrait être représenté par cinq fonctions de base (deux fonctions de type s et trois fonctions de type p).

Les électrons de valence jouent un rôle prédominant dans l'ensemble des processus chimiques d'intérêt. Il est donc normal de vouloir raffiner la description orbitale de ces électrons. Pour ce faire, il est possible de représenter les orbitales de valence non pas par une gaussienne, mais par une combinaison linéaire de plusieurs gaussiennes. Cette description permet d'ajuster la densité électronique en fonction de l'environnement chimique. La nomenclature usuelle pour définir de telles bases est la suivante :

$$A - \underbrace{BCD} G \quad (2.9)$$

où A représente le nombre de gaussiennes composant chaque fonction de base des électrons de cœur. B et C signifient que chaque orbitale de valence est décrite par deux fonctions de base (ou trois dans le cas où D est défini), chacune consistant en une combinaison linéaire de B et C (et D) gaussiennes. Dans cet exemple, on parle d'une base zeta-double (ou zeta-triple si D est défini). G signifie que les primitives sont des gaussiennes.

Il est possible d'ajouter des fonctions de polarisation aux fonctions de base. Ces fonctions, dénotées par \* (appliqué à tous les atomes sauf hydrogène et hélium) ou \*\* (appliqué à tous les atomes) sont des fonctions supplémentaires possédant un nœud supplémentaire. Par exemple, dans le cas d'un atome d'hydrogène, la fonction de polarisation ajoutée serait de type p. Cet ajout permet de considérer une certaine flexibilité en permettant de déformer la symétrie autour des noyaux atomiques pour permettre une représentation plus fidèle des liens chimiques.

Finalement, il est possible d'ajouter des fonctions diffuses aux bases, dénotées par + (appliqué à tous les atomes sauf hydrogène et hélium) et ++ (appliqué à tous les atomes). Ces fonctions permettent de mieux définir les orbitales loin de noyaux, où la densité est généralement mal définie. Cet ajout est important notamment dans la simulation des anions, ou dans les systèmes faisant intervenir des interactions à longue distance.

Dans toutes les expériences réalisées, des fonctions primitives de type gaussien ont été utilisées. Toutefois, étant donné la nature différente des molécules, des bases différentes ont été utilisées. Dans le cas de l'observation des états de transition (section 2.4), la base 6-31++G\*\* a été utilisée. Ce choix se base sur le fait que tous les atomes en cause appartiennent exclusivement à la première et la deuxième période du tableau périodique et qu'ils n'ont par conséquent pas d'orbitales d remplies partiellement ou totalement. Il a par ailleurs été montré<sup>35</sup> que la base 6-31G permet de simuler de manière adéquate les composés de cet espace chimique. La base 6-311++G\*\*<sup>36</sup> a quant à elle été utilisée pour calculer les énergies d'excitation des molécules photoactives (section 2.3). Cette décision a été motivée par le fait que les molécules d'intérêt sont constituées d'atomes de la troisième période, possédant donc des orbitales d.

### 2.1.2. TD-DFT

Nous avons vu que la DFT nous permet d'obtenir l'énergie du système en fonction de la densité électronique. Les théorèmes de Hohenberg-Kohn ont permis de confirmer la validité mathématique de ce postulat (section 1.5.4). Cette validité mathématique n'est toutefois valable que pour les états fondamentaux. Cette limitation ne pose généralement pas de problème considérant que nous étudions dans la vaste majorité des cas l'état fondamental d'un système moléculaire. Toutefois, lorsque nous souhaitons observer les effets d'une perturbation sur un système, notamment ceux d'un champ électromagnétique, la DFT n'est pas, *a priori*, la méthode adéquate pour déterminer certaines quantités d'intérêt. Nous devons dès lors penser à utiliser la théorie de la fonctionnelle de la densité dépendante du temps (TD-DFT), qui est l'extension de la DFT nécessaire à l'évaluation des propriétés de systèmes moléculaires dans des états excités. Initialement formulée par le théorème de Runge et Gross<sup>37</sup>, elle permet d'observer les effets d'une perturbation sur un système quantique. Nous discuterons plus en détail de la méthodologie utilisée pour calculer le spectre d'absorption dans la section 2.2.4.

### 2.1.3. Choix de GAUSSIAN

GAUSSIAN 03<sup>38</sup> a été utilisé pour tous les calculs. Ce choix a été motivé par le fait que les fonctionnelles et méthodes sont implémentées dans ce logiciel. GAUSSIAN est aussi accessible sur les infrastructures à notre disposition.

## 2.2. Méthodologie

Une méthodologie a été développée pour simuler tant les interactions intramoléculaires des mésomères que les énergies d'excitation des molécules photoactives. Les différentes composantes de cette méthodologie sont expliquées dans les sous-sections suivantes avec leurs particularités propres.

### 2.2.1. Construction des structures

Toutes les structures ont été générées à l'aide du logiciel Avogadro<sup>39</sup>. Une optimisation rapide de la géométrie a été réalisée à même ce logiciel. L'obtention de coordonnées internes compatibles avec GAUSSIAN a été possible par le module d'exportation d'Avogadro.

### 2.2.2. Détermination des états de transition

Les algorithmes d'optimisation de géométrie passant par la détection de minima locaux sont assez bien connus. L'évaluation des états de transition est par contre plus complexe à faire puisque, contrairement à une optimisation de géométrie pour laquelle nous souhaitons optimiser tous les paramètres (par exemple, la longueur de liaison et l'angle), l'évaluation d'un état de transition passe par la détermination d'un seul paramètre maximisant l'énergie, puis par l'optimisation ultérieure de tous les autres paramètres afin de la minimiser.

Pour arriver à ce résultat, il est possible d'utiliser la méthode STQN<sup>40</sup> (Synchronous Transit-Guided Quasi-Newton) de GAUSSIAN. Cette méthode consiste à considérer des géométries atomiques initiale ( $\vec{R}$ ) et finale ( $\vec{P}$ ), qui sont les géométries optimisées du réactif et du produit, et à réaliser une interpolation quadratique entre ces deux vecteurs. En déplaçant un vecteur  $\vec{X}$  sur le chemin de l'interpolation, on peut déterminer une géométrie qui correspondra à un maximum sur la surface de potentiel. Concrètement, dans ce chapitre, étant donné que notre réactif et notre produit sont géométriquement très similaires et que seul les coordonnées d'un atome change de manière significative, réaliser cette interpolation consiste à chercher un vecteur  $\vec{X}$  qui déplacera le proton de sorte à ce qu'il soit distribué de part et d'autre des atomes donneurs et accepteurs (Figure 2.11). Par la suite, il ne reste qu'à minimiser les autres coordonnées géométriques par la méthode des vecteurs propres pour obtenir l'état de transition.

### 2.2.3. Observation des orbitales naturelles

Les orbitales telles que définies précédemment dans ce chapitre et en introduction sont dites canoniques. Elles sont en général délocalisées et n'ont pas de lien clair avec la notion de liaison chimique localisée. Par exemple, les orbitales canoniques de valence du méthane s'expriment par une combinaison linéaire des orbitales atomiques du carbone et des hydrogènes (Figure 2.2).

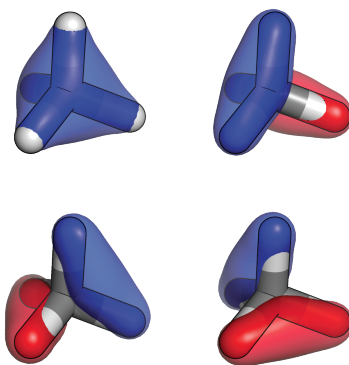


Figure 2.2 – Orbitales canoniques de valence du méthane.

Cette représentation n'aide pas forcément à avoir une idée sur les liens chimiques impliqués dans le méthane. Pour obtenir une meilleure description des liens chimiques tels que nous les concevons, il est possible d'appliquer une transformation orthogonale à un sous-ensemble d'orbitales moléculaires<sup>41</sup>. Le résultat permet d'obtenir une description plus intuitive des interactions moléculaires (Figure 2.3).

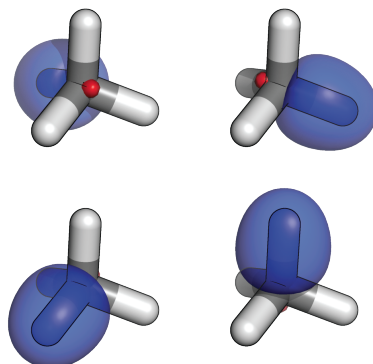


Figure 2.3 – Orbitales naturelles de valence du méthane.

Bien que la fonction d'onde totale reste invariable à cette transformation, le calcul des orbitales naturelles entraîne toutefois la perte des énergies orbitales individuelles. Les orbitales naturelles seront utilisées dans la section 2.4.1 pour évaluer le changement d'hybridation atomique.

#### 2.2.4. Détermination du spectre d'absorption par TD-DFT

Le calcul du spectre d'absorption optique peut être réalisé en utilisant la théorie de la réponse linéaire<sup>42</sup>. De manière générale, il s'agit de trouver la réponse linéaire  $\delta\langle\hat{a}\rangle(t)$  pour une propriété  $a$  à une perturbation. Dans le cas qui nous intéresse, la réponse observée est la susceptibilité générale ( $\chi$ ). Pour obtenir la réponse, une transformée de Fourier est réalisée, puis, en supposant que le système est initialement dans son état fondamental, il est possible de résoudre par la théorie des perturbations. La représentation de la susceptibilité générale dans le spectre de fréquence est donnée par<sup>42</sup>:

$$\chi(\vec{r}, \vec{r}', \omega) = \sum_n \left( \frac{\langle \psi_0 | \hat{a}^\dagger(\vec{r}) \hat{a}(\vec{r}) | \psi_n \rangle \langle \psi_n | \hat{a}^\dagger(\vec{r}') \hat{a}(\vec{r}') | \psi_0 \rangle}{\omega - (E_n - E_0) + i\eta} - \frac{\langle \psi_0 | \hat{a}^\dagger(\vec{r}') \hat{a}(\vec{r}') | \psi_n \rangle \langle \psi_n | \hat{a}^\dagger(\vec{r}) \hat{a}(\vec{r}) | \psi_0 \rangle}{\omega + (E_n - E_0) + i\eta} \right) \quad (2.10)$$

où  $\psi_0$ ,  $\psi_n$  et  $E_0$ ,  $E_n$  sont respectivement les fonctions d'onde et les énergies à l'état fondamental et à l'état excité,  $\hat{a}^\dagger(\vec{r})\hat{a}(\vec{r})$  est l'opérateur de champ,  $\omega$  est la fréquence et le paramètre  $\eta$  est un réel positif. À partir de cette équation, il est possible d'obtenir la polarisabilité dynamique :

$$\alpha(\omega) = \sum_n \frac{f_n}{(E_n - E_0)^2 - \omega^2} \quad (2.11)$$

où

$$f_n = \frac{2}{3} (E_n - E_0) (|\langle \psi_0 | \hat{x} | \psi_n \rangle|^2 + |\langle \psi_0 | \hat{y} | \psi_n \rangle|^2 + |\langle \psi_0 | \hat{z} | \psi_n \rangle|^2) \quad (2.12)$$

où  $\hat{x}$ ,  $\hat{y}$  et  $\hat{z}$  sont les opérateurs de position. Ces dernières équations permettent d'obtenir les énergies d'excitation (équation 2.11) et les forces d'oscillateur (équation 2.12). La TD-DFT sera utilisée dans la section suivante pour déterminer les énergies d'excitation de molécules photoactives.

### 2.3. Détermination des énergies d'excitation de molécules photoactives

Les cellules photovoltaïques organiques constituent une avenue intéressante de conversion d'énergie solaire et une alternative potentielle aux cellules inorganiques en raison de leur faible coût potentiel et de la possibilité de créer de nouveaux matériaux flexibles ayant la capacité de capter l'énergie environnante<sup>43</sup>. Ces dispositifs sont habituellement constitués de deux électrodes séparées par une hétérojonction (couche active) contenant un mélange de donneur et d'accepteur d'électron (Figure 2.4)<sup>44</sup>.

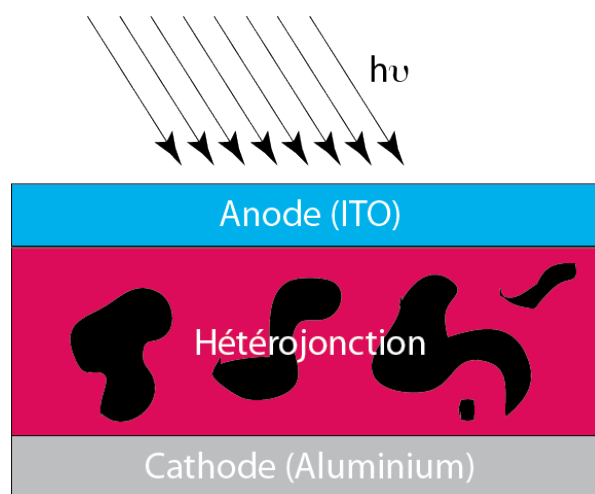


Figure 2.4 – Schéma d'une cellule photovoltaïque organique.

L'anode consiste typiquement en une couche d'oxyde d'indium-étain (ITO) qui possède la particularité d'être à la fois conducteur électrique et transparent en couches minces. La cathode est normalement composée d'un métal, qui, en plus de jouer le rôle de conducteur électrique, permet d'agir à titre d'agent réfléchissant. Finalement, la couche active est composée au minimum d'un donneur et d'un accepteur d'électron, les deux composés étant des molécules organiques dans le cadre de cette étude.



La première étape dans la conception d'une cellule photovoltaïque efficace consiste de choisir de bons donneurs et accepteurs. Essentiellement, un photon peut exciter un électron de l'orbitale HOMO du donneur vers son orbitale LUMO. L'électron peut par la suite être transféré à l'orbitale LUMO de l'accepteur et collecté à l'électrode générant ainsi un photocourant (Figure 2.5).

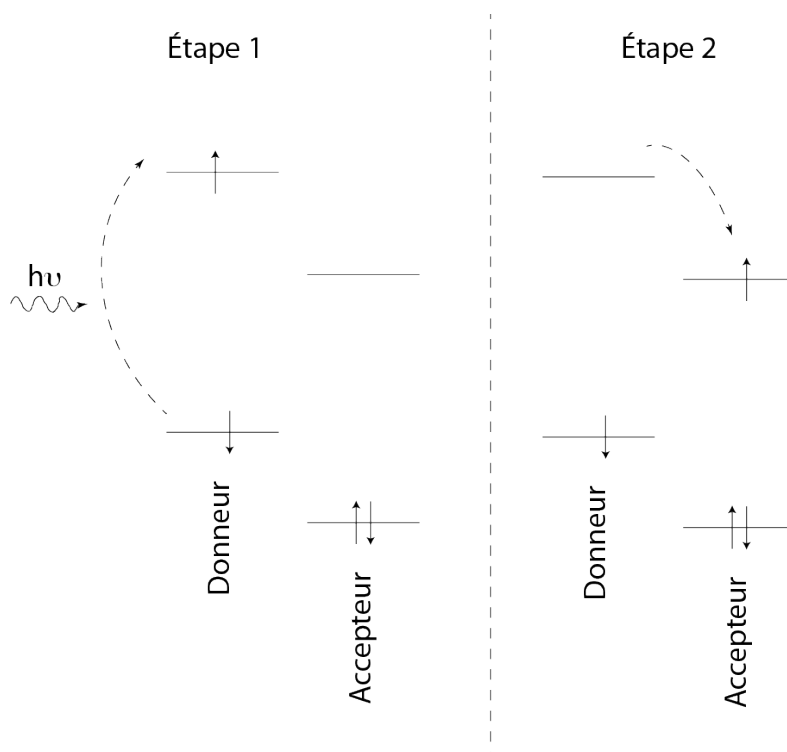


Figure 2.5 –Fonctionnement de la couche active.

Les orbitales HOMO et LUMO du donneur et de l'accepteur sont donc des variables critiques dans la conception de nouvelles couches actives. Dans un premier temps, la différence entre l'orbitale LUMO et HOMO du donneur doit correspondre à une énergie radiative importante du spectre lumineux. Par ailleurs, l'orbitale LUMO du donneur doit être plus élevée que celle de l'accepteur.

En plus de la complémentarité des orbitales du donneur et de l'accepteur, ceux-ci doivent être assez compatibles pour favoriser une morphologie maximisant l'échange électronique entre les deux espèces.

Dans cette section, nous allons voir comment les énergies d'excitation et les orbitales HOMO et LUMO des molécules **2.1** à **2.4** (Figure 2.6) ont été calculées. Nous comparerons ces énergies avec les résultats expérimentaux obtenus pour l'un des candidats. L'idée derrière la conception de ces composés était d'utiliser des cœurs de type indénofluorène (composés **2.1** et **2.2**) et de type indacénodithiophène (composés **2.3** et **2.4**) comme modulateurs d'orbitales HOMO et des bras composés de thiophènes substitués par des groupements de types dicyanovinyles comme modulateurs d'orbitales LUMO.

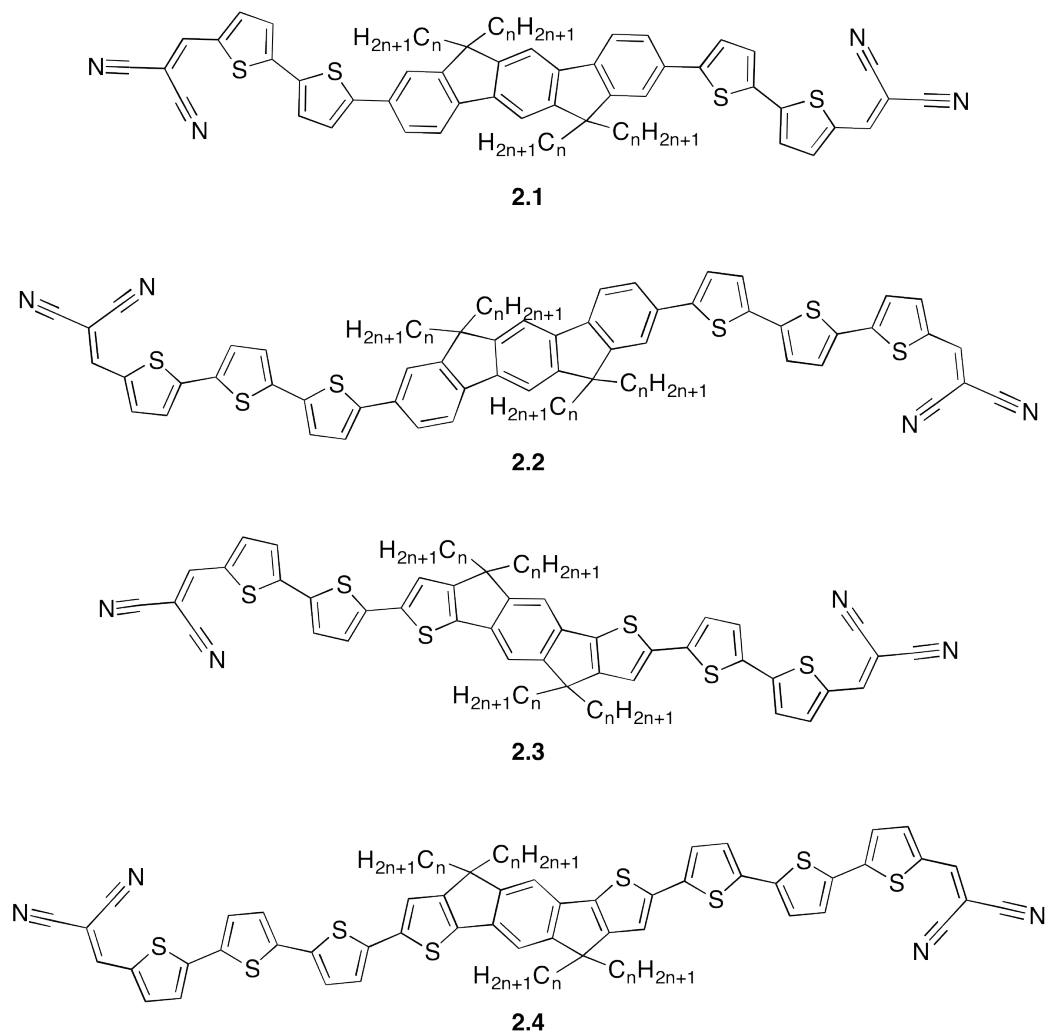


Figure 2.6 – Molécules à l'étude.

### 2.3.1. Sommaire des résultats expérimentaux disponibles

Le composé **2.1** ( $n = 8$ ) a été synthétisé par Alice Heskia<sup>†</sup> et Pierre-Louis Brunner<sup>‡</sup>. Les énergies des orbitales HOMO et LUMO ont été déduites à partir des potentiels d'oxydation

<sup>†</sup> Département de Chimie, Université de Montréal, Montréal, Québec H3C 3J7, Canada

<sup>‡</sup> Solaris Chem Inc. Montréal, Québec H3T 2B1, Canada

obtenus par voltampérométrie cyclique. Des valeurs de -5,54 eV et -2,92 eV ont été obtenues pour les orbitales HOMO et LUMO respectivement, pour une énergie d'excitation de 2,62 eV. Ces résultats expérimentaux ont permis d'évaluer la qualité des calculs et les performances de la fonctionnelle PBE0.

### 2.3.2. Résultats computationnels

Pour obtenir les orbitales et les énergies d'excitation des molécules **2.1** à **2.4** ( $n = 1$ ), les structures ont été modélisées et optimisées selon la méthodologie décrite plus tôt. Les géométries obtenues diffèrent principalement en raison de la nature du cœur de chaque molécule. En effet, les structures **2.1** et **2.2** possèdent un cœur de type indénofluorène ce qui empêche la planarité du corps aromatique (Figure 2.7).

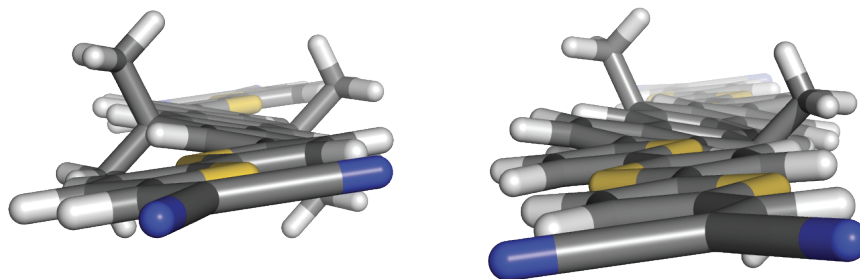


Figure 2.7 – Géométrie optimisée de la structure **2.1** (à gauche) et **2.2** (à droite). Un angle d'environ 15,5° est observé.

Cette rupture dans la planarité est attribuable à l'encombrement stérique causé par la proximité des protons entre le cœur et les extrémités des composés **2.1** et **2.2** (Figure 2.8).

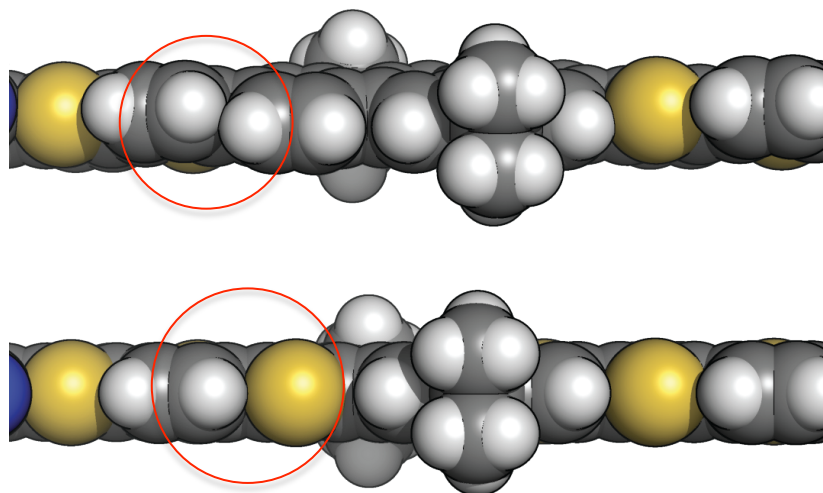


Figure 2.8 – Encombrement stérique entre le cœur et les extrémités du composé **2.1**. Une comparaison est faite avec le composé **2.3**.

En contrepartie, l'optimisation des structures **2.3** et **2.4**, possédant un cœur de type indacénodithiophène conduit à l'obtention d'une structure totalement plane (Figure 2.9).

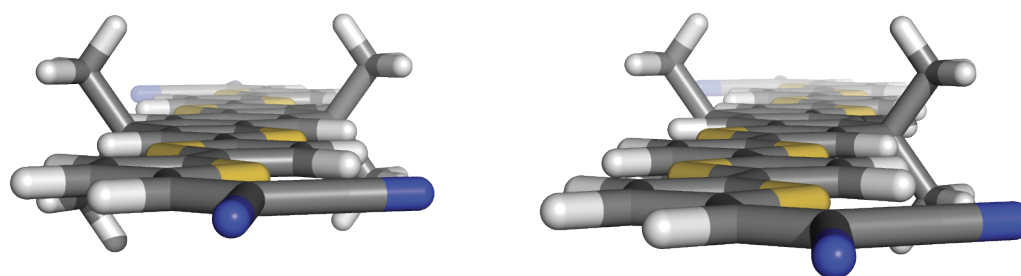


Figure 2.9 – Géométrie optimisée de la structure **2.3** (à gauche) et **2.4** (à droite).

Ces géométries ont été par la suite utilisées pour déterminer les énergies d'excitation les plus probables et les orbitales HOMO et LUMO de tous les composés. Comme prévu, on observe les orbitales HOMO de tous les composés au niveau des cœurs et les orbitales

LUMO au niveau des bras (Figure 2.10). Finalement, la modélisation par TD-DFT a permis d'obtenir les énergies d'excitation des composés (Tableau I).

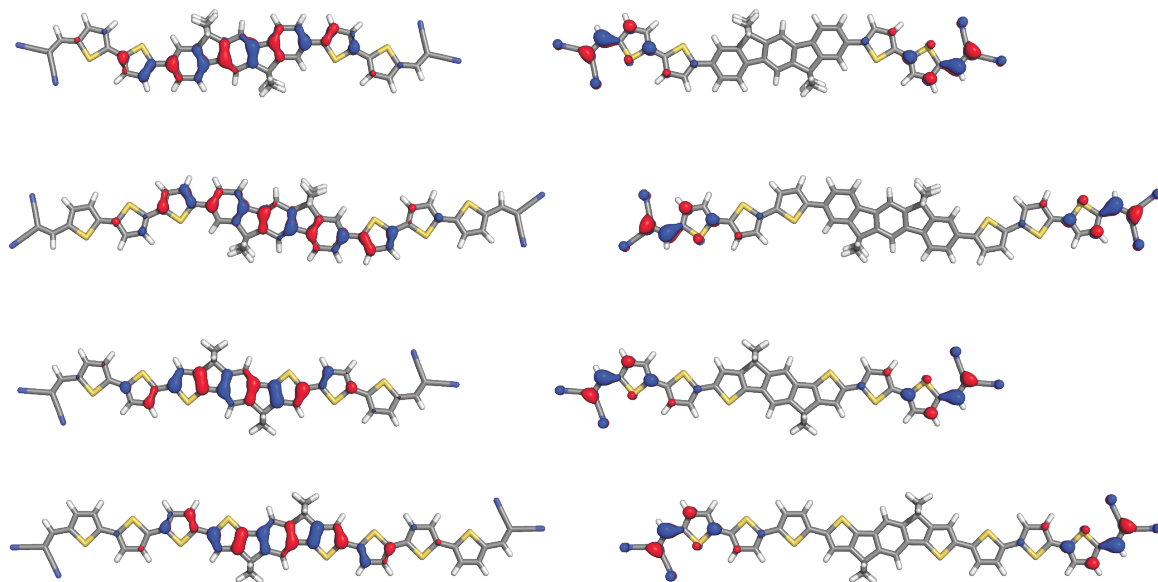


Figure 2.10 – Orbitales HOMO (à gauche) et LUMO (à droite) des composés **2.1** à **2.4** respectivement.

Tableau I – Énergies d'excitation expérimentales ( $\lambda_{exp}$ ) et calculées ( $\lambda$ ) et forces d'oscillateur (F) des molécules **2.1** à **2.4**.

Composé	$\lambda_{exp}$ (nm)	$\lambda$ (nm)	F (U.A.)
<b>2.1</b>	473	480,4	2,08
<b>2.2</b>	-	506,8	2,00
<b>2.3</b>	-	532,2	1,93
<b>2.4</b>	-	549,6	1,74

On peut constater que les énergies d'excitation expérimentales et calculées pour le composé **2.1** diffèrent de moins de 5 %, ce qui suggère que la méthode employée permet de rendre de manière adéquate le portrait électronique des composés suggérés. Une comparaison systématique des composés **2.2** – **2.4** n'a pu toutefois être réalisée étant donné que la méthode de synthèse de ces molécules reste encore à développer.

Par ailleurs, on constate que le changement du cœur indénofluorène pour un cœur indacénodithiophène contribue à diminuer l'énergie d'excitation. Ce changement est principalement attribuable au fait que l'angle entre le cœur et les bras (Figure 2.7) contribue à limiter la délocalisation<sup>45</sup> sur l'ensemble du squelette aromatique pour les composés **2.1** et **2.2**. De plus, on constate que l'énergie d'excitation diminue à mesure que le nombre d'unités thiophène augmente. Pour les mêmes raisons, l'ajout d'unités contribue à augmenter la conjugaison sur l'ensemble du composé<sup>46</sup>.

### 2.3.3. Conclusions

En somme, les molécules **2.1** à **2.4** ont été modélisées et leur énergie d'excitation a été évaluée par TD-DFT. De manière générale, on constate que la délocalisation contribue à diminuer la différence d'énergie entre l'orbitale HOMO et LUMO. La méthodologie a été validée en comparant les résultats computationnels aux résultats expérimentaux disponibles. Une erreur de moins de 5 % a été observée en comparant les deux types de résultats. Ces résultats nous ont encouragés à réutiliser la même fonctionnelle conjointement à la DFT pour évaluer les états de transitions d'analogues du 1,3,5-triazapentadiène.

### 2.4. Observation des interactions intramoléculaires par l'étude des états de transition de molécules analogues au 1,3,5-triazapentadiène

Le pont hydrogène intervient dans une vaste proportion de questions fondamentales en chimie. Ce type d'interaction, mis en évidence dès 1912<sup>47</sup> pour rationaliser la basicité relative de l'hydroxyde de tetraméthylammonium par rapport à l'hydroxyde de triméthylammonium, est reconnu surtout pour son rôle comme élément structurant de multiples processus biologiques et chimiques. Bien qu'historiquement défini comme une interaction de nature majoritairement électrostatique, le pont hydrogène est considéré aujourd'hui comme étant une interaction complexe de partage de proton, où il existe une modularité induite par les propriétés acides-bases du donneur et de l'accepteur<sup>48</sup>.

Les structures analogues aux 1,3-dicétones ont été largement étudiées afin de mieux comprendre et expliquer la nature du pont hydrogène<sup>48-49</sup>. L'étude des états de transition et du chemin réactionnel entre les réactifs et les produits de ce type de structures a permis de faire un lien clair entre la forme du chemin et la force du pont hydrogène<sup>49c</sup>. Dans cette



section, nous utiliserons une approche similaire pour étudier les ponts hydrogène des fragments analogues au 1,3,5-triazapentadiène (Figure 2.11).

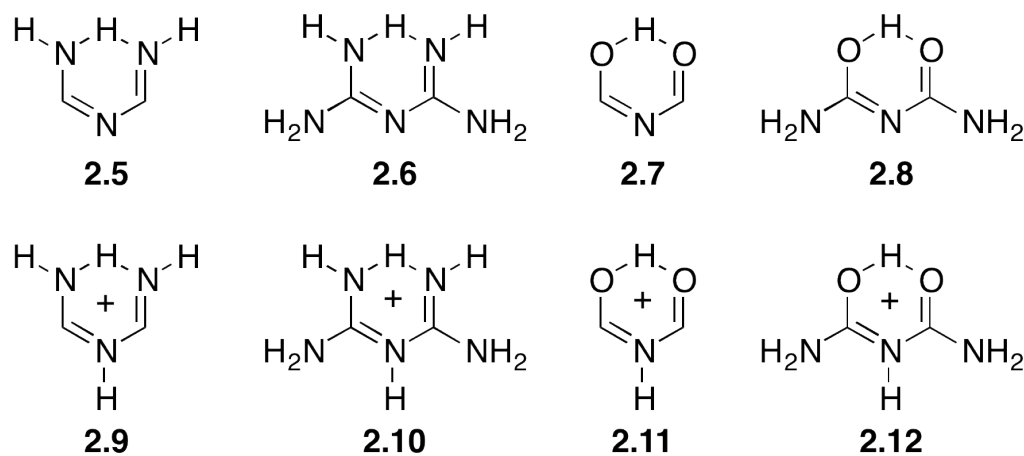


Figure 2.11 – Fragments à l'étude

Pour ce faire, nous observerons les états de transition entre les mésomères en concentrant notre analyse sur trois aspects modulaires de ces molécules : (1) la charge, (2) la nature des atomes donneurs et accepteurs et (3) la présence de substituants. Cette analyse nous permettra de mieux comprendre les structures et les propriétés associatives de molécules plus complexes telles que la bis-diaminotriazine **2.13**.

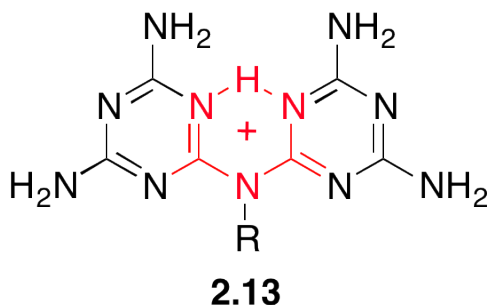


Figure 2.12 – Motif de reconnaissance bis-diaminotriazine **2.13**. Le centre, commun aux molécules **2.5-2.12**, est dessiné en rouge.

#### 2.4.1. Résultats computationnels

Pour comprendre les différents effets sur la stabilité des états de transition des molécules analogues au 1,3,5-triazapentadiène, les molécules **2.5** à **2.12** ont été modélisées et les états de transition ont été calculés. La différence d'énergie entre les états de transition et l'état fondamental ( $\Delta E_{Tr}$ ) a par la suite été calculée. Une comparaison des  $\Delta E_{Tr}$  des différentes molécules ( $\Delta \Delta E_{Tr}$ ) a été réalisée de manière à voir l'effet de la charge, de la nature des atomes donneurs et accepteurs et de la présence des substituants de manière indépendante. Le Tableau II résume les résultats obtenus :

Tableau II – Sommaire des différences d'énergies entre l'état de transition et l'état fondamental ( $\Delta E_{Tr}$ ) et de la comparaison de ces différences ( $\Delta\Delta E_{Tr}$ ).

Effet	N° Molécule A	N° Molécule B	$\Delta E_{Tr}$ A (kJ/mol)	$\Delta E_{Tr}$ B (kJ/mol)	$\Delta\Delta E_{Tr}$ (kJ/mol)
Charge	<b>2.5</b>	<b>2.9</b>	22,00	19,28	<b>-2,724</b>
	<b>2.6</b>	<b>2.10</b>	16,82	11,54	<b>-5,271</b>
	<b>2.7</b>	<b>2.11</b>	8,357	9,638	<b>1,281</b>
	<b>2.8</b>	<b>2.12</b>	2,934	2,123	<b>-0,811</b>
Nature des atomes donneurs et accepteurs	<b>2.7</b>	<b>2.5</b>	8,357	22,00	<b>13,65</b>
	<b>2.11</b>	<b>2.9</b>	9,638	19,28	<b>9,642</b>
	<b>2.8</b>	<b>2.6</b>	2,934	16,82	<b>13,88</b>
	<b>2.12</b>	<b>2.10</b>	2,123	11,54	<b>9,422</b>
Présence de substituants	<b>2.5</b>	<b>2.6</b>	22,00	16,82	<b>-5,188</b>
	<b>2.9</b>	<b>2.10</b>	19,28	11,54	<b>-7,735</b>
	<b>2.7</b>	<b>2.8</b>	8,357	2,934	<b>-5,423</b>
	<b>2.11</b>	<b>2.12</b>	9,638	2,123	<b>-7,515</b>

En considérant les  $\Delta\Delta E_{Tr}$  groupés par effet, on peut voir que, de manière générale, la présence d'une charge positive stabilise légèrement l'état de transition, la substitution des oxygènes par des azotes pour les donneurs et accepteurs déstabilise l'état de transition et finalement la présence de substituants amino stabilise l'état de transition. Afin de rationaliser le lien entre la stabilisation de l'état de transition et l'existence d'un pont hydrogène plus fort à l'état fondamental, les orbitales naturelles ont été observées. La Figure 2.13 montre que pour la structure ayant la valeur  $\Delta E_{Tr}$  la plus élevée (**2.5**), la paire libre sur l'atome accepteur ne contribue pas à une liaison de type pont-hydrogène alors que

pour la structure correspondant à l'état de transition la plus stabilisée (**2.12**), la paire libre contribue de manière importante et qu'un lien partiel est formé.

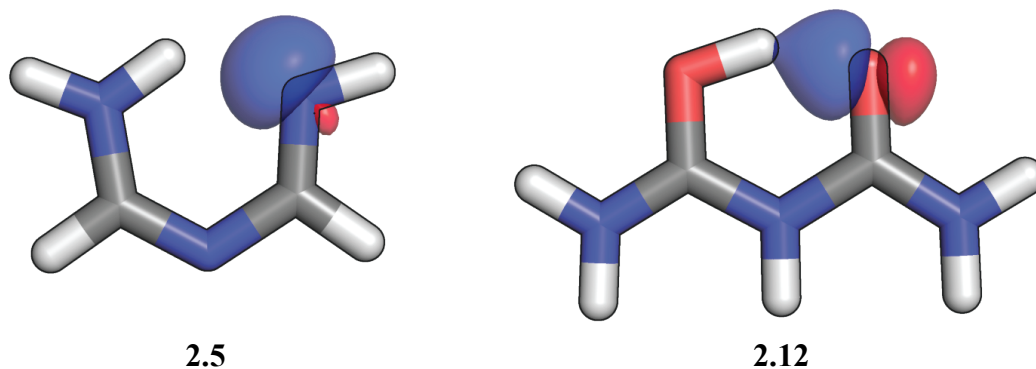


Figure 2.13 – Orbitale naturelle de la paire libre des molécules **2.5** et **2.12**. La valeur de  $\Delta E_{Tr}$  pour les structures est de 22,00 kJ/mol et 2,123 kJ/mol respectivement.

Pour confirmer cette relation, la variation de distance proton-accepteur a été calculée et comparée au  $\Delta\Delta E_{Tr}$ .

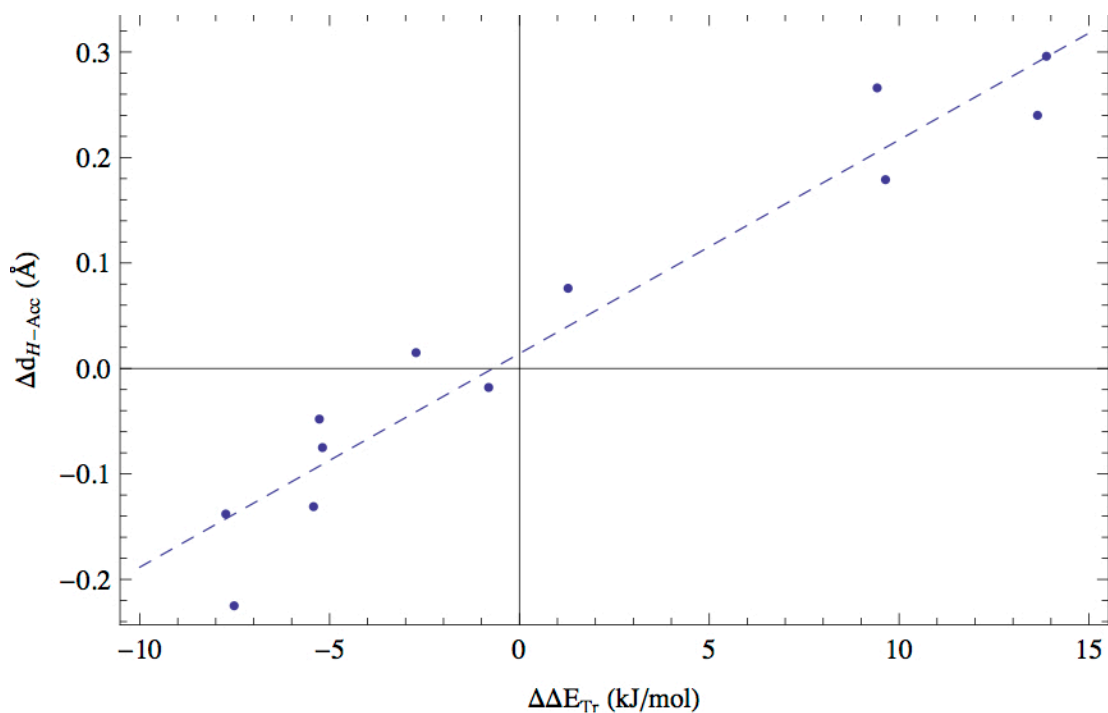
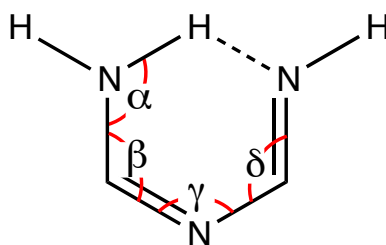


Figure 2.14 – Relation entre la variation de distance proton-accepteur ( $\Delta d_{H-Acc}$ ) et la différence d'énergie de l'état de transition ( $\Delta\Delta E_{Tr}$ ). Un coefficient de détermination de 0,93 a été obtenu pour la relation  $\Delta d_{H-Acc} = 2,025 \times 10^{-2} \times \Delta\Delta E_{Tr} - 1,413 \times 10^{-2}$ .

La Figure 2.14 montre qu'il existe une relation entre la distance proton-accepteur et la stabilisation de l'état de transition. Plus particulièrement, elle indique que l'état de transition est stabilisé par un proton plus également partagé entre le donneur et l'accepteur à l'état fondamental. Afin de montrer quels paramètres géométriques favorisent ce meilleur partage du proton entre l'accepteur et le donneur, les angles  $\alpha$ ,  $\beta$ ,  $\gamma$  et  $\delta$  (Figure 2.15) ont été calculés pour toutes les structures.



## 2.5

Figure 2.15 – Angles  $\alpha$ ,  $\beta$ ,  $\gamma$  et  $\delta$  pour la molécule **2.5**.

Afin de déduire une relation globale, un angle combiné, consistant en la somme des angles  $\alpha$ ,  $\beta$ ,  $\gamma$  et  $\delta$  a été calculé et la variation de cet angle a été comparée au  $\Delta\Delta E_{\text{Tr}}$  (Figure 2.16). On constate que la stabilisation de l'état de transition est étroitement liée à la combinaison des angles. Pour un angle combiné réduit, on observe une stabilisation de l'état de transition.

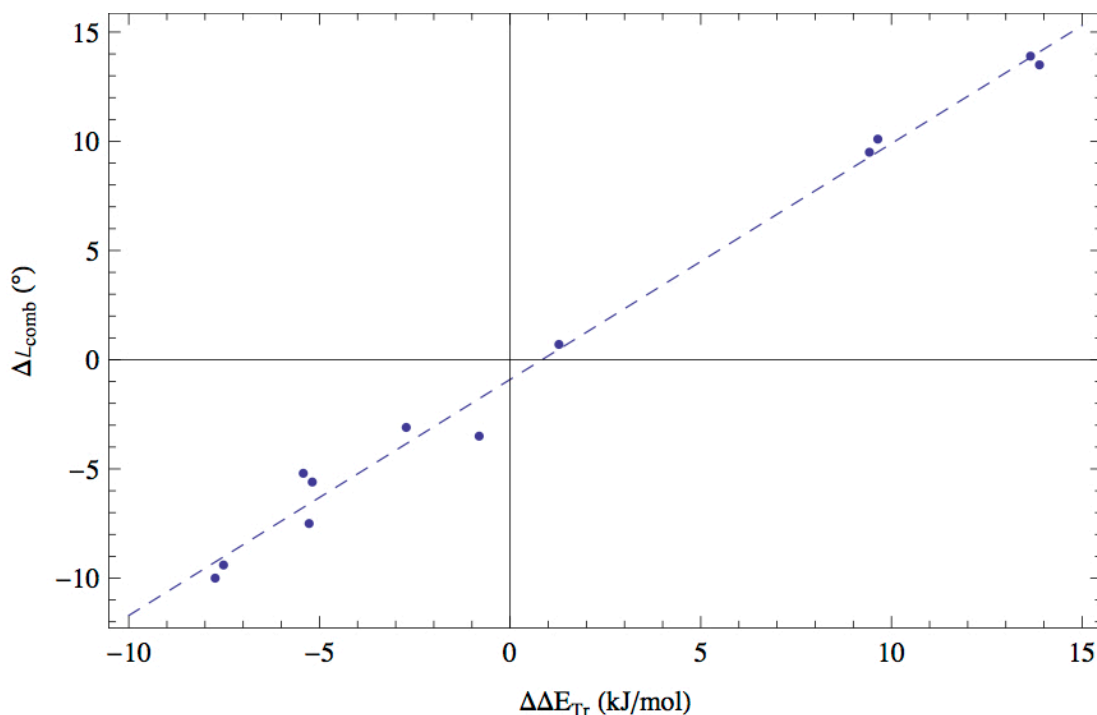


Figure 2.16 – Relation entre la variation de l'angle combiné ( $\Delta\angle_{comb}$ ) et la différence d'énergie de l'état de transition ( $\Delta\Delta E_{Tr}$ ). Un coefficient de détermination de 0,99 a été obtenu pour la relation  $\Delta\angle_{comb} = 1,081 \times \Delta\Delta E_{Tr} - 9,058 \times 10^{-1}$ .

Ces derniers résultats nous ont encouragés à étudier individuellement les angles  $\alpha$ ,  $\beta$ ,  $\gamma$  et  $\delta$  afin de comprendre la contribution de chaque effet dans la stabilisation ou la déstabilisation de l'état de transition. D'un point de vue électronique, l'angle formé entre trois atomes est défini par la nature et la contribution des orbitales atomiques dans la formation des liens interatomiques. La variation d'hybridation ( $\Delta hyb$ ) des atomes au centre des angles a donc été observée pour chaque effet (Tableau III). Celle-ci a été calculée en comparant le caractère p des atomes centraux.

Tableau III – Variation d’hybridation des atomes centraux aux angles  $\alpha$ ,  $\beta$ ,  $\gamma$  et  $\delta$  pour l’effet de charge, l’effet de la nature des atomes donneurs et accepteurs et l’effet de substituants.

Effet	N° Molécule A	N° Molécule B	$\Delta\text{hyb}_\alpha$	$\Delta\text{hyb}_\beta$	$\Delta\text{hyb}_\gamma$	$\Delta\text{hyb}_\delta$
Charge	<b>2.5</b>	<b>2.9</b>	0,065	0,190	-0,065	0,215
	<b>2.6</b>	<b>2.10</b>	-0,160	0,065	0,055	0,105
	<b>2.7</b>	<b>2.11</b>	-0,060	0,110	0,010	0,160
	<b>2.8</b>	<b>2.12</b>	-0,040	0,135	0,015	0,250
Nature des atomes donneurs et accepteurs	<b>2.7</b>	<b>2.5</b>	-0,560	-0,125	-0,105	-0,040
	<b>2.11</b>	<b>2.9</b>	-0,435	-0,045	-0,180	0,015
	<b>2.8</b>	<b>2.6</b>	-0,315	-0,020	-0,070	0,075
	<b>2.12</b>	<b>2.10</b>	-0,435	-0,050	-0,030	-0,070
Présence de substituants	<b>2.5</b>	<b>2.6</b>	0,280	0,175	0,050	0,150
	<b>2.9</b>	<b>2.10</b>	0,055	0,050	0,170	0,040
	<b>2.7</b>	<b>2.8</b>	0,035	0,030	0,015	0,035
	<b>2.11</b>	<b>2.12</b>	0,055	0,055	0,020	0,125

De manière générale, un changement d’hybridation induisant un caractère p moins important (diminution du  $\Delta\text{hyb}$ ) entraîne une augmentation de l’angle formé par deux liaisons sur l’atome étudié. Cette relation illustre notamment le lien entre les géométries linéaire, trigonale planaire et tétraédrique des hybridations  $\text{sp}$ ,  $\text{sp}^2$  et  $\text{sp}^3$  des atomes. Dans le cadre de cette analyse, nous considérerons donc qu’un  $\Delta\text{hyb}$  positif contribuera à stabiliser l’état de transition puisqu’il diminue l’angle formé par deux liens sur l’atome à



l'étude. Étant donné la similarité des différents cas, seule la comparaison entre les composés les plus simples sera considérée.

### 1. Effet de charge

Dans un premier temps, nous pouvons voir que de manière générale, la variation de l'hybridation  $\Delta\text{hyb}_\alpha$  est légèrement négative, ce qui contribue à déstabiliser l'état de transition, mais qu'en contrepartie, les variations  $\Delta\text{hyb}_\beta$  et  $\Delta\text{hyb}_\delta$  sont positives dans une plus grande mesure. Globalement, on observe donc que l'hybridation peut contribuer à favoriser une stabilisation de l'état de transition. Ce changement d'hybridation est explicable en observant les orbitales  $\pi$  des composés.

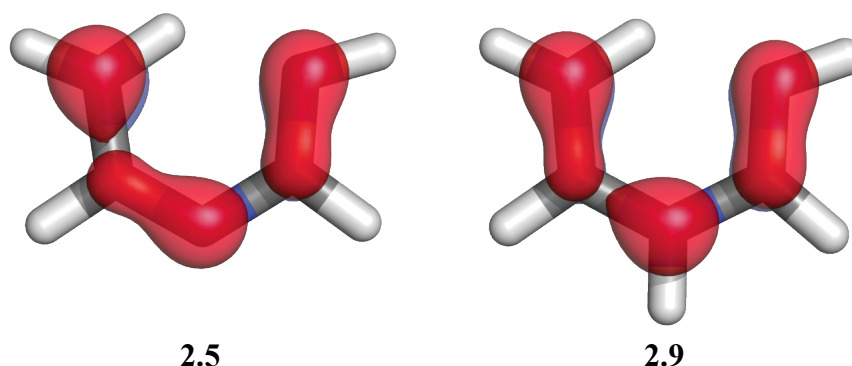


Figure 2.17 – Orbitales naturelles  $\pi$  du composé **2.5** (à gauche) et **2.9** (à droite).

On constate dans la Figure 2.17 que la présence de la charge renforce le caractère double de la liaison entre le donneur (atome  $\alpha$ ) et le carbone (atome  $\beta$ ) tout en diminuant le caractère double entre ce carbone et l'azote central (atome  $\gamma$ ). La donation provenant de l'atome  $\alpha$ , il est donc normal que le caractère p de l'orbitale  $\sigma$  de l'atome  $\alpha$  diminue. Par conséquent, le

lien  $\pi$  entre le carbone  $\beta$  et l'azote central  $\gamma$  étant partiellement brisé, on constate que l'hybridation du lien  $\sigma$  augmente.

## 2. Effet de la nature des atomes donneurs et accepteurs

Le changement des atomes d'oxygène pour des atomes d'azote comme donneurs et accepteurs contribue à diminuer l'hybridation de tous les atomes étudiés. Plus particulièrement, nous pouvons remarquer que la variation de l'hybridation du donneur est particulièrement négative. Ce comportement est dû au fait que l'azote est moins électronégatif que l'oxygène et, par conséquent, plus électrodonneur. Le lien double partiel entre l'azote  $\alpha$  et le carbone  $\beta$  est donc plus fort et le lien  $\sigma$  possède par conséquent un caractère p moins important (Figure 2.18).

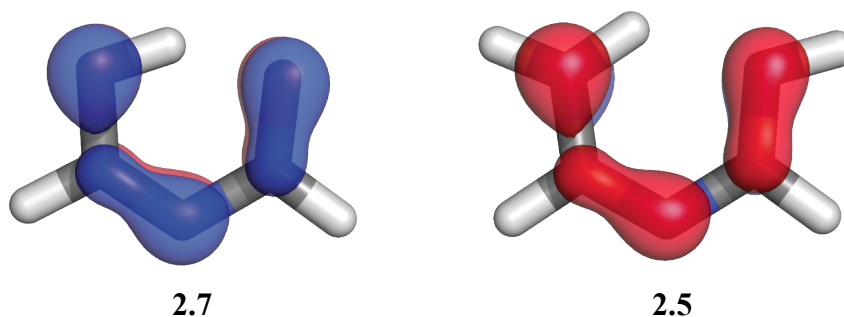


Figure 2.18 – Orbitales naturelles  $\pi$  du composé **2.7** (à gauche) et **2.5** (à droite).

## 3. Effet de la présence de groupements substituants

Finalement, la présence des amines comme groupements substituants contribue à augmenter l'hybridation de tous les atomes étudiés et donc à stabiliser l'état de transition. Le caractère électrodonneur des groupements substituants amino contribue à diminuer

partiellement la donation des autres atomes donneurs sur l'ensemble du composé ce qui contribue globalement à augmenter l'hybridation de tous les atomes (Figure 2.19).

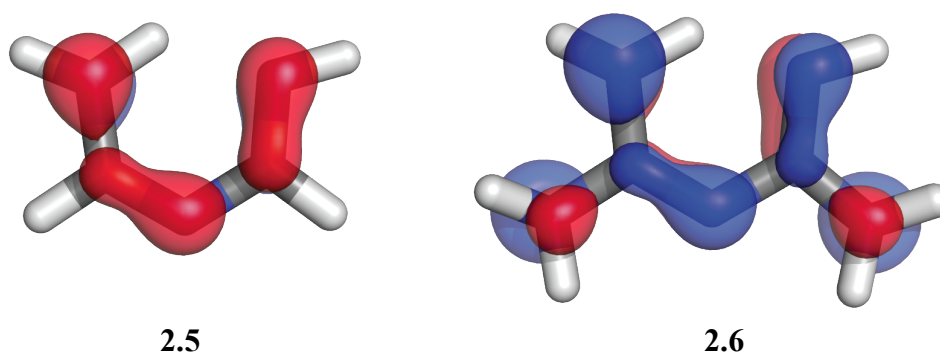


Figure 2.19 – Orbitales naturelles  $\pi$  du composé **2.5** (à gauche) et **2.6** (à droite).

#### 2.4.2. Limites et conclusions

En somme, les interactions intramoléculaires de composés analogues au 1,3,5-triazapentadiène ont été observées en étudiant les états de transition entre les deux mésomères des structures **2.5** à **2.12**. Trois effets ont été considérés et traités de manière indépendante : l'effet de charge, l'effet de la nature des atomes donneurs et accepteurs et l'effet de groupements substituants.

De manière générale, il a été montré que la présence de la charge stabilise légèrement l'état de transition, que la substitution des atomes d'oxygène pour des atomes d'azote comme donneurs et accepteurs déstabilise l'état de transition et que la présence de substituants amino stabilise l'état de transition.

Il a été établi qu'indépendamment de l'effet étudié, l'état de transition est stabilisé lorsque la distance entre le proton et l'accepteur diminue. L'étude de l'angle combiné formé des angles  $\alpha$ ,  $\beta$ ,  $\gamma$  et  $\delta$  a été réalisée et il a été montré qu'il existe une corrélation étroite entre la diminution de cet angle et la stabilisation de l'état de transition.

Pour lier chaque effet à une stabilisation ou une déstabilisation de l'état de transition, l'analyse du changement d'hybridation des structures sans et avec l'effet a été réalisée. Il a été montré que de manière générale, la présence de la charge diminue le caractère p de l'orbitale  $\sigma$  du donneur (atome  $\alpha$ ), tout en augmentant le caractère p de l'orbitale  $\sigma$  des atomes des carbones ( $\beta$  et  $\delta$ ). Le changement des atomes d'oxygène pour des atomes d'azote comme donneurs et accepteurs contribue à diminuer l'hybridation de tous les atomes étudiés. Finalement, la présence de substituants amino contribue à augmenter globalement l'hybridation de tous les atomes étudiés.

L'observation des orbitales naturelles s'est révélée une méthode très utile pour déduire et comprendre l'importance des effets observés. Il est toutefois important de considérer que cette étude se voulait restreinte à une meilleure compréhension de la stabilisation de l'état de transition de la bis-diaminotriazine **2.13**. Dans une perspective plus large de prédiction des états de transition de molécules quelconque, cette étude n'est bien entendu pas suffisante, mais la méthodologie développée pourrait servir de base solide à l'analyse systématique d'une banque de composés plus étendue.

## Chapitre 3: Utilisation de la dynamique moléculaire dans l'étude des interactions supramoléculaires

L'observation des interactions supramoléculaires implique la simulation des interactions évoluant à une échelle nanométrique. Cette réalité limite donc, dans la majorité des cas, l'utilisation des méthodes quantiques précédemment discutées.

Dans ce chapitre, nous discuterons des simplifications qui peuvent être faites pour outrepasser les difficultés computationnelles imposées par la taille des systèmes moléculaires. Nous appliquerons ces simplifications à deux systèmes supramoléculaires : d'une part, nous observerons les processus d'encapsulation de pyrènes dans des trimères dérivés d'acides choliques, puis nous simulerons le comportement d'un rotaxane-parapluie dans divers solvants et à l'interface d'un système biphasique en proposant certaines étapes importantes du changement de phase. Ces systèmes complexes seront en mesure de montrer comment les modèles utilisés représentent de façon adéquate le subtil équilibre entre les différents types d'interactions discutés dans ce chapitre.

Comme nous l'avons vu en introduction, la mécanique moléculaire vient résoudre l'impasse de la complexité mathématique imposée par les méthodes *ab initio*. L'utilisation adéquate de ce type de méthodes se base sur trois principes fondamentaux<sup>50</sup> : l'hypothèse que le système tend vers un état thermodynamiquement favorable, qu'il existe une notion d'additivité des potentiels d'interaction et finalement que ces potentiels peuvent être transférés d'un groupe de molécules vers un autre.

L'application du premier principe en chimie computationnelle est illustrée notamment par le paradoxe de Levinthal<sup>51</sup>. En 1969, Cyrus Levinthal met en lumière le fait qu'il est mathématiquement impossible qu'une protéine puisse se replier en une conformation thermodynamiquement stable en explorant systématiquement chaque conformation possible. L'exemple de Levinthal peut très bien s'appliquer aux situations de cette étude,

considérant que nous étudions ici des systèmes supramoléculaires possédant un très grand nombre de conformations possibles.

Le deuxième principe stipule que les potentiels obtenus en paramétrant de petites simulations moléculaires sur des résultats expérimentaux peuvent être réutilisés pour simuler des systèmes plus étendus qui possèdent des groupements moléculaires similaires et représentatifs. Ce principe se base sur l'observation qu'en conditions similaires, les longueurs et les angles de liaisons de certains groupements fonctionnels se ressemblent. Il est toutefois important de noter que certaines contraintes chimiques tendent à changer ces valeurs de référence. Par exemple, la délocalisation électronique peut induire des changements dans les longueurs ou les angles de liaisons<sup>52</sup>. Ce problème peut être contourné notamment en donnant un caractère spécial à certains atomes. Ce principe constitue d'ailleurs la raison principale pour laquelle il est primordial de faire un choix judicieux lorsqu'un champ de force est utilisé. En effet, il n'existe pas un champ de force adéquat à toute situation rencontrée. Les champs de force sont paramétrés sur diverses expériences afin de reproduire par simulation les résultats de celles-ci. Si les systèmes étudiés lors du paramétrage du modèle sont très différents du système que nous souhaitons modéliser, il y a de fortes probabilités que les paramètres ne soient pas très bien adaptés au contexte qui nous intéresse et que la simulation ne donne pas de résultats justes.

Finalement, le dernier principe stipule qu'il existe une notion d'additivité des potentiels. Ce principe permet de séparer les contributions de l'énergie moléculaire en potentiels associés à divers observables physiques ou à des modèles décrivant les systèmes moléculaires. Par exemple, dans la majorité des cas, les champs de force contiennent les informations permettant de décrire des interactions interatomiques liantes (liens covalents) ou non liantes (van der Waals, électrostatiques et d'autres types). Cette additivité permet de séparer ces contributions, ce qui simplifie les calculs et permet de dégager des algorithmes de résolution puissants, mais qui limite toutefois la modélisation à des systèmes où le concept

de réarrangement atomique n'existe pas. En effet, étant donné que le type des interactions est spécifiquement décidé à l'avance, il n'est pas possible de changer ou d'observer une modification de la nature de celle-ci au cours de la simulation. La notion de réaction chimique n'est donc pas applicable à ce type de simulation.

### **3.1. Mécanique statistique**

Lorsqu'il est question d'explorer les propriétés macroscopiques d'un système à partir de simulations microscopiques, il est nécessaire d'utiliser la mécanique statistique pour établir une connexion entre ces deux notions. La mécanique statistique est l'étude des systèmes macroscopiques d'un point de vue moléculaire<sup>53</sup>. Pour ce faire, des événements (par exemple des conformations) sont générés et une quantité statistique (par exemple une grandeur thermodynamique) est obtenue en moyennant la valeur calculée pour chaque événement.

Une série d'événements possédant des états microscopiques différents, mais un état macroscopique identique est appelé un ensemble. On appelle un ensemble canonique, un ensemble de systèmes clos qui possèdent la même composition, le même volume et la même température (ensemble NVT) et qui peuvent échanger leur énergie<sup>54</sup>. L'échange d'énergie est généralement réalisé en utilisant un thermostat (section 3.4.6). Afin de se rapprocher des cas concrets comme ceux dont nous discuterons dans ce chapitre, il est possible de considérer l'ensemble isotherme-isobare. Cet ensemble contient des systèmes clos qui possèdent la même composition, la même température et la même pression (ensemble NPT) et qui peuvent donc échanger énergie et volume. L'utilisation d'un thermostat et d'un barostat (section 3.4.7) est dès lors nécessaire.

Il existe principalement deux méthodologies pour générer des ensembles : l'approche stochastique et l'approche déterministe. L'approche stochastique, plus connue sous le nom

de Monte Carlo, consiste à explorer la surface d'énergie potentielle en modifiant la géométrie du système de manière plus ou moins aléatoire. Les principales étapes sont les suivantes<sup>55</sup> :

1. Génération d'une géométrie initiale  $\mathbf{X}_0^N$ .
2. Génération d'une nouvelle géométrie  $\mathbf{X}_a^N$  de manière aléatoire.
3. Calcul de la probabilité de transition  $W(0, a)$ .
4. Génération d'un nombre aléatoire  $0 \leq r \leq 1$ .
5. Si  $W(0, a) < r$ ,  $\mathbf{X}_0^N$  est conservée comme telle, sinon elle est remplacée par  $\mathbf{X}_a^N$ .
6. Passage à l'étape 2.

L'approche déterministe, appliquée en dynamique moléculaire, consiste quant à elle observer l'évolution temporelle d'un système et permet d'obtenir des trajectoires. C'est d'ailleurs pour cette raison que la dynamique moléculaire a été choisie dans le cadre des simulations de ce chapitre. Une description plus étendue de la dynamique moléculaire est faite à la section 3.3.

Finalement, il est possible d'évaluer des propriétés thermodynamiques en utilisant principalement trois quantités : la fonction de partition ( $Z$ ), la distribution de Boltzmann ( $P(\vec{p}, \vec{q})$ ) et la valeur moyenne de l'ensemble pour une quantité  $A$  voulue ( $\langle A \rangle$ ). Celles-ci prennent les formes suivantes :

$$Z = \iint e^{\frac{-H(\vec{p}, \vec{q})}{kT}} d\vec{p} d\vec{q} \quad (3.1)$$



$$P(\vec{p}, \vec{q}) = \frac{1}{Z} e^{\frac{-H(\vec{p}, \vec{q})}{kT}} \quad (3.2)$$

$$\langle A \rangle = \iint A(\vec{p}, \vec{q}) P(\vec{p}, \vec{q}) d\vec{p} d\vec{q} \quad (3.3)$$

où  $\vec{p}$  et  $\vec{q}$  sont respectivement les coordonnées et les vitesses du système,  $H(\vec{p}, \vec{q})$  est l'Hamiltonien du système,  $k$  est la constante de Boltzmann et  $T$  est la température du système.

### 3.2. Choix du champ de force

Tel que nous l'avons vu, certains principes forment la base du choix de l'utilisation de la mécanique moléculaire dans l'étude de systèmes. Ces mêmes principes peuvent nous guider quant à la décision de favoriser un champ de force plutôt qu'un autre. En introduction, nous avons discuté de la forme fonctionnelle que prenaient certains champs de force. Nous verrons dans cette section comment les principaux champs de force précédemment présentés sont paramétrés afin de nous aider à valider notre choix final.

Répondant historiquement à des besoins de modélisation de systèmes biologiques, les champs de force sont aujourd'hui très bien paramétrés pour les protéines, les acides nucléiques et leurs dérivés. Il est toutefois plus rare et plus difficile de trouver des outils adaptés à l'étude des plus petites molécules qui foisonnent le monde de la chimie. Parmi les champs de force disponibles, MMFF94, CGENFF, OPLS et GAFF se démarquent par le fait qu'ils sont conçus pour simuler des molécules organiques en solution.

### 3.2.1. MMFF94

MMFF94<sup>56</sup> (Merck Molecular Force Field) est un champ de force spécifiquement conçu pour simuler des ligands interagissant avec des protéines. Développé dans les laboratoires de recherche de Merck, il a l'avantage d'être bien paramétré pour les molécules organiques, particulièrement en milieu aqueux<sup>56-57</sup>. Par ailleurs, étant donné la diversité des composés auxquels il s'adresse, les paramètres de MMFF94 découlent de calculs *ab initio* et non pas d'observations expérimentales.

La liste suivante regroupe les principales classes des composés incluses dans les paramètres de MMFF94 :

- Les alcanes, alcènes, composés aromatiques et composés conjugués.
- Les cycles à 4 ou 5 membres.
- Les hétérocycles à 5 ou 6 membres.
- Les alcools, phénols et éthers.
- Les amines et imines
- Les amines hydroxylées et N-oxydées.
- Les amides, dipeptides, urées, imides et acides hydroxamiques.
- Les aldéhydes et cétones.
- Les cétales, acétales, hemicétales et hemiacétales,
- Les acides carboxyliques et les esters.
- Les  $\beta$ -cétoacides et  $\beta$ -cétoesters.
- Les thiols et sulfides.
- Les halogénures.
- Les amines, imines, amidines, guanidines, pyridines et imidazoles chargées.
- Les carboxylates

### 3.2.2. CGENFF (CHARMM)

CHARMM<sup>58</sup> (Chemistry at HARvard Molecular Mechanics) est un ensemble de champs de force qui était à l'origine paramétré pour la simulation des acides nucléiques et des acides aminés. Avec le temps, la nécessité de travailler avec des molécules organiques diverses a encouragé le développement de paramètres plus généraux. Plus particulièrement, CGENFF<sup>57</sup> a été conçu pour simuler les petites molécules organiques qui interagissent avec les molécules biologiques. Les molécules types de CGENFF se classent en deux catégories : les composés hétérocycliques et les groupements fonctionnels de base. La stratégie derrière la construction des paramètres a été d'utiliser les mêmes composés utilisés pour construire les champs de force des protéines et acides nucléiques. Par exemple, le phénol a été paramétré comme précurseur de la tyrosine (Figure 3.1).

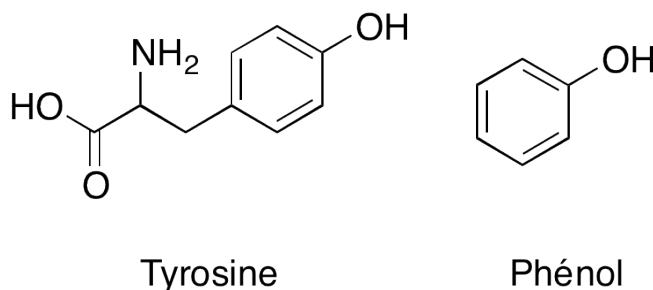


Figure 3.1 – Illustration de la tyrosine et du phénol.

Les paramètres de Lennard-Jones ont été obtenus en comparant les densités liquides et les enthalpies de vaporisation calculées avec les valeurs expérimentales disponibles.

Finalement, afin d'étendre les paramètres, 68 composés contenant des hétérocycles ont été ajoutés. Au total, 445 composés modèles ont été paramétrés, consistant notamment en les classes suivantes :

- Les nitrobenzènes

- Les biphényles
- Les opioïdes
- Les alcynes
- Les hydrazines
- Les sulfoxydes
- Les acides biliaires
- Les halogénures aliphatiques et aromatiques
- Les indolizines
- Les amidines
- Les urées
- Les cétones

### 3.2.3. OPLS

OPLS<sup>59</sup> (Optimized Potentials for Liquid Simulations) est un champ de force développé spécifiquement pour la simulation des molécules organiques et des peptides en phase liquide. Les paramètres des torsions et des interactions non liantes ont été obtenus par calculs *ab initio* sur plus de 50 molécules organiques neutres et ioniques. Les paramètres des liens et des angles proviennent du champ de force AMBER (sous-section 3.2.4). Les classes de composés auxquelles OPLS fait référence sont notamment :

- Les alcanes et alcènes
- Les alcools et éthers
- Les acétals
- Les thiols, sulfides et disulfides
- Les aldéhydes et cétones
- Les amides

Finalement, des simulations Monte-Carlo<sup>55</sup> de plus de 30 liquides organiques purs ont été réalisées et comparées aux propriétés thermodynamiques disponibles afin d'améliorer les paramètres en milieu liquide.

#### 3.2.4. GAFF (AMBER)

GAFF<sup>60</sup> (Generalized AMBER Force Field) fait partie de l'ensemble des champs de force AMBER<sup>61</sup>. Il a été spécifiquement développé pour la simulation de molécules organiques, principalement pour faire l'analyse et la conception de pharmacophores. Ses paramètres reposent sur une optimisation MP2/6-31G\* de plus de 3000 composés et sur des calculs simple-point MP4/6-311G(d,p) de 1260 molécules. En sa version actuelle, il représente l'espace chimique des atomes d'hydrogène, de carbone, d'azote, de soufre, de phosphore et des halogènes fluor, chlore, brome et iode. Cet espace est représenté par 55 types d'atomes.

Étant donné que GAFF est paramétré sur des composés similaires à ceux de cette étude (deuxième principe), il se plaçait déjà comme un candidat de choix pour cette étude. Certaines autres caractéristiques, en accord avec le troisième principe, sont venues renforcer cette décision :

- Une étude<sup>60</sup> de 74 composés en phase gazeuse a montré que les géométries obtenues par GAFF et comparées aux structures cristallines sont similaires aux résultats obtenus avec Tripos 5.2<sup>62</sup> et meilleures que ceux calculés avec MMF94 et CHARMM. De plus, la déviation moyenne de longueurs et d'angles de liaison est comparable aux valeurs obtenues avec MMFF94 et meilleure que celles calculées par Tripos 5.2 et DREIDING<sup>63</sup>, deux champs de force génériques utilisés pour des simulations moléculaires.

- Cette même étude a montré qu'une comparaison des ponts hydrogène de 22 paires de bases entre une optimisation MP2/6-31G\* et GAFF donne une différence de longueur du pont de 0,255 Å pour une différence d'énergie moyenne de 1,21 kcal/mol.
- Afin d'observer la reproduction des énergies relatives entre un paire de conformères, la différence d'énergie entre 71 paires de conformères a été réalisée et comparée avec les résultats expérimentaux. La différence entre les résultats computationnels et expérimentaux est d'environ 0.50 kcal/mol. Le faible écart obtenu avec GAFF montre à quel point l'additivité des potentiels est applicable aux systèmes organiques.

Par ailleurs, le calcul des masses volumiques et des enthalpies de vaporisation du méthanol, du *N*-méthylacétamide et du butane, des composés couvrant un domaine de polarité large, a été réalisé en utilisant MMFF94, OPLS et GAFF<sup>64</sup>. La comparaison avec les valeurs de la littérature montre des erreurs systématiquement plus grandes avec MMFF94 (21 % à 30 %) qu'avec OPLS (1 % à 2 %) et GAFF (4 % à 5 %)<sup>64</sup>.

Finalement, le choix s'est arrêté sur GAFF étant donné qu'il était possible d'automatiser son utilisation pour nos besoins en utilisant l'outil Antechamber<sup>65</sup> et qu'il était compatible avec GROMACS, le logiciel de dynamique moléculaire que nous voulions utiliser.

### 3.3. Choix de la dynamique moléculaire, de GROMACS et d'Antechamber

Tel que mentionné précédemment, la dynamique moléculaire est l'étude de l'évolution de systèmes moléculaires dans le temps. Elle utilise la méthode des différences finies pour générer des trajectoires. Concrètement, il s'agit de résoudre l'équation classique du mouvement sur de petits incréments de temps à partir de coordonnées initiales :

$$m_i \frac{d^2 \vec{x}_i}{dt^2} = \vec{f}_i \quad \vec{f}_i = - \frac{dU}{d\vec{x}_i} \quad (3.4)$$

Les forces  $\vec{f}_i$  sont obtenues par les potentiels empiriques décrits par le champ de force. Le calcul de celles-ci découle de la dérivation de l'équation de l'énergie potentielle et cinétique.

La dernière étape consistait à choisir le logiciel de dynamique. Le choix s'est porté sur GROMACS<sup>66</sup>, un logiciel de dynamique moléculaire développé à l'Université de Groningen et utilisé notamment par les chimistes et biochimistes de Folding@home<sup>67</sup>, un projet international visant à étudier des systèmes biologiques de grande envergure. GROMACS a été privilégié pour tous les calculs en raison de sa grande modularité, de l'implémentation de tous les algorithmes qui seront décrits dans la prochaine section, de la parallélisation de son code qui permet des simulations extrêmement rapide et pour sa capacité à pouvoir facilement travailler avec des systèmes moléculaires organiques. Antechamber a été utilisé pour générer automatiquement les nombreux fichiers de travail nécessaires à la dynamique.

### 3.4. Méthodologie

Une méthodologie générale a été développée pour simuler tant l'encapsulation de pyrènes par un trimère d'acide cholique que les interactions intermoléculaires du rotaxane-parapluie à l'interface d'une séparation de phase entre l'eau et le dichlorométhane. Les différentes composantes de cette méthodologie sont expliquées dans les sous-sections suivantes.

### 3.4.1. Construction des structures et calcul des charges atomiques

Les structures ont toutes été générées à l'aide du logiciel Avogadro<sup>67</sup>. À partir des structures, la géométrie a été optimisée et les charges atomiques ont été calculées avec MOPAC2009<sup>68</sup>. Le modèle semi-empirique RM1<sup>69</sup> a été choisi pour toutes les optimisations. Il s'agit d'une version améliorée de AM1<sup>10</sup> avec de nouveaux paramètres pour les atomes d'hydrogène, carbone, azote, oxygène, phosphore, soufre, fluor, chlore, brome et iode. Cette méthode donne généralement, par rapport aux méthodes typiques (AM1, PM3<sup>70</sup> et PM5<sup>70</sup>), des géométries, des moments dipolaires et des chaleurs de formation comparables ou meilleures que ces modèles<sup>69</sup>. Les charges partielles des atomes sont utilisées pour calculer le potentiel électrostatique (équation 1.20).

### 3.4.2. Solvation

Le solvant joue un rôle prédominant dans les diverses simulations de ce chapitre. Afin de minimiser les effets de surface, la méthode de la solvation périodique (PBC) a été utilisée. Cette méthode permet de réinsérer une molécule qui ressortirait d'un côté de la boîte à l'extrémité opposée de celle-ci. La Figure 3.2 schématise la réinsertion.

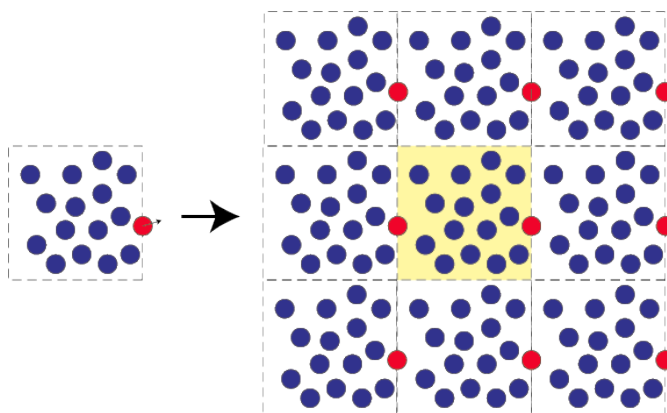


Figure 3.2 – Représentation d'une boîte de solvant par la méthode PBC. Une particule sortant de la boîte est illustrée en rouge (à gauche). La figure de droite montre comment cette particule est remise dans la boîte par le côté opposé (encadré jaune à droite).



### 3.4.3. Traitement des interactions non liantes

De toutes les interactions considérées dans l'équation 1.16, c'est le nombre de termes coulombiques et de van der Waals qui croît le plus rapidement et qui devient assez vite pénible à calculer. En effet, étant donné la double sommation nécessaire au calcul de ce type d'interactions, la complexité de ce terme évolue en  $O(n^2)$  par rapport au nombre  $n$  d'atomes du système. Deux simplifications ont été mises en place afin d'accélérer les calculs.

D'une part, la contribution des interactions de van der Waals a été tronquée à partir d'une certaine distance. En effet, en observant le terme du potentiel de Lennard-Jones dans l'équation 1.21, on peut voir qu'il existe une composante répulsive et une composante attractive, toutes deux normalisées par la distance entre les deux particules mise à la puissance 12 et 6 respectivement. Il est donc facile de voir que ces contributions convergent rapidement vers 0 à mesure que la distance entre les particules augmente. On considérera donc une distance au-delà de laquelle le potentiel sera simplement nul.

D'autre part, la contribution des interactions électrostatiques a été traitée selon la méthode PME<sup>71</sup> développée d'après la méthode originale de Paul Peter Ewald. Celle-ci décrit le potentiel d'interaction comme la somme de deux termes :

$$\varphi(r) = \varphi_{\text{courte distance}}(r) + \varphi_{\text{longue distance}}(r) \quad (3.5)$$

La contribution à courte distance  $\varphi_{\text{courte distance}}$  est définie selon le terme direct spécifié dans l'équation 1.20. À longue portée, les densités de charge sont plutôt évaluées sur une maille

discrète, interpolées par des B-splines et traitées par une transformée de Fourier rapide. Cette méthodologie permet d'obtenir une complexité en  $O(n \times \log(n))$ <sup>71</sup>.

Pour toutes les interactions non liantes, une distance de 10 Å est considérée comme la valeur limite de changement de méthode.

#### 3.4.4. Minimisation d'énergie

Sitôt les systèmes construits et les boîtes solvatées, l'énergie des systèmes devait être initialement minimisée pour éviter les faux contacts entre les molécules. Deux étapes de minimisation ont été systématiquement réalisées avant chaque dynamique :

1. Tout d'abord, une optimisation rapide de la géométrie du système a été réalisée avec l'algorithme *steepest-descent*<sup>72</sup> implémenté dans GROMACS. Un maximum de 5000 étapes était accepté, après quoi le calcul s'arrêtait automatiquement. Cette première étape permettait de converger rapidement vers un minimum local.
2. Ensuite, la méthode des gradients conjugués<sup>73</sup> a été utilisée. Cette étape supplémentaire permettait de raffiner la géométrie sans toutefois obtenir le comportement oscillatoire occasionné par la recherche *steepest-descent*<sup>50</sup>.

#### 3.4.5. Algorithme d'intégration *Leap-Frog*

L'algorithme d'intégration *Leap-Frog*<sup>74</sup> a été utilisé pour toutes les dynamiques. L'idée derrière *Leap-Frog* est d'utiliser les positions calculées pour générer des vitesses et de réutiliser ces vitesses pour calculer de nouvelles positions. Les positions et les vitesses sont calculées comme suit :

$$\vec{x}_{i+1} = \vec{x}_i + \vec{v}_{i+1/2} \delta t \quad (3.6)$$

$$\vec{v}_{i+3/2} = \vec{v}_{i+1/2} + \vec{f}(x_{i+1}) \delta t \quad (3.7)$$

Les positions  $\vec{x}_i$  ( $i = 0$ ) sont données par les coordonnées initiales des structures, résultat des minimisations d'énergie. La seule information manquante pour démarrer l'algorithme concerne les vitesses à  $\vec{v}_{i+1/2}$  ( $i = 0$ ). Ces vitesses sont calculées dans GROMACS en utilisant une distribution de Maxwell-Boltzmann à la température spécifiée (équation 3.8) :

$$p(\vec{v}_i) = \sqrt{\frac{m_i}{2\pi kT}} e^{\left[ \frac{-m_i \vec{v}_i^2}{2kT} \right]} \quad (3.8)$$

où  $p(\vec{v}_i)$  représente la probabilité d'obtenir une vitesse  $\vec{v}_i$  pour la particule  $i$ ,  $m_i$ , la masse de la particule,  $k$  la constante de Boltzmann et  $T$  la température.

#### 3.4.6. Thermostat de Berendsen

Les dynamiques moléculaires sont des systèmes qui évoluent dans le temps. Afin de nous assurer que les conformations générées correspondent à une représentation statistique adéquate de la population que nous souhaitons étudier, nous devons contrôler certains paramètres. C'est pourquoi nous utilisons les thermostats et les barostats.

Le thermostat de Berendsen<sup>75</sup> permet de contrôler la température du système par l'utilisation d'un module externe. L'effet principal est que la température est corrigée

graduellement au fur et à mesure de la dynamique et que les fluctuations de l'énergie cinétique sont supprimées. Cette méthodologie ne génère donc pas un ensemble canonique, mais l'erreur varie selon  $1/N$  (où  $N$  représente le nombre d'atomes). Pour les systèmes étendus, l'erreur est donc négligeable<sup>76</sup>. Le système est appelé un ensemble NVT, car le nombre de molécules ( $N$ ), le volume ( $V$ ) et la température ( $T$ ) sont constants. La Figure 3.3 montre l'effet du thermostat sur la température d'une dynamique NVT.

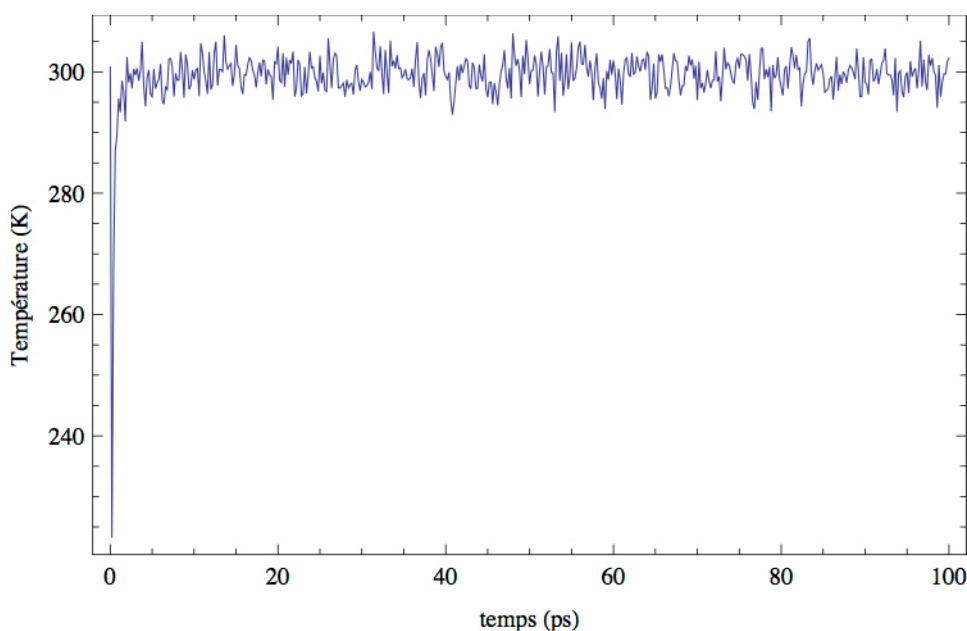


Figure 3.3 – Effet du thermostat de Berendsen sur la température d'un système aqueux. La température est rapidement ajustée et conservée à 300 K.

#### 3.4.7. Barostat de Parrinello-Rahman

Le second paramètre que nous voulons contrôler est la pression. Pour ce faire, nous utilisons le barostat de Parrinello-Rahman<sup>77</sup>. Ce barostat permet de modifier la taille de la boîte de solvant en fonction des fluctuations de pression. L'équation 3.9 définit comment la taille de la boîte varie :

$$\frac{d\mathbf{b}^2}{dt^2} = V\mathbf{W}^{-1}\mathbf{b}^{T^{-1}}(\mathbf{P} - \mathbf{P}_{ref}) \quad (3.9)$$

où  $V$  correspond au volume de la boîte,  $\mathbf{b}$  à une représentation matricielle de la boîte,  $\mathbf{P}$  et  $\mathbf{P}_{ref}$  aux matrices de la pression et de la pression de référence respectivement et  $\mathbf{W}$  une matrice permettant de déterminer la force du couplage. Cette matrice est générée notamment en prenant la compressibilité isotherme (Tableau IV) comme paramètre.

$$(\mathbf{W}^{-1})_{ij} = \frac{4\pi^2\beta_{ij}}{3\tau_p^2L} \quad (3.10)$$

où  $\beta_{ij}$  est la compressibilité isotherme,  $\tau_p$  une constante de couplage de la pression et  $L$ , le plus grand élément de  $\mathbf{b}$ . La compressibilité isotherme est une mesure du changement relatif d'un volume d'un fluide ou d'un liquide face à un changement de pression à température constante. Dans le cas des systèmes anisotropes,  $\beta_{ij}$  est généralement une matrice diagonale définie comme étant :

$$\beta_{i=j} = -\frac{1}{V} \left( \frac{\partial V}{\partial p} \right)_T \quad \beta_{i \neq j} = 0 \quad (3.11)$$

Les systèmes obtenus par ce barostat sont appelés systèmes NPT, car le nombre de molécules (N), la pression (P) et la température (T) sont constants. La Figure 3.4 montre l'effet du barostat de Parinello-Rahman sur la masse volumique d'une boîte d'eau.

L'observation des masses volumiques des systèmes est une validation implicite des paramètres et du champ de force utilisé.

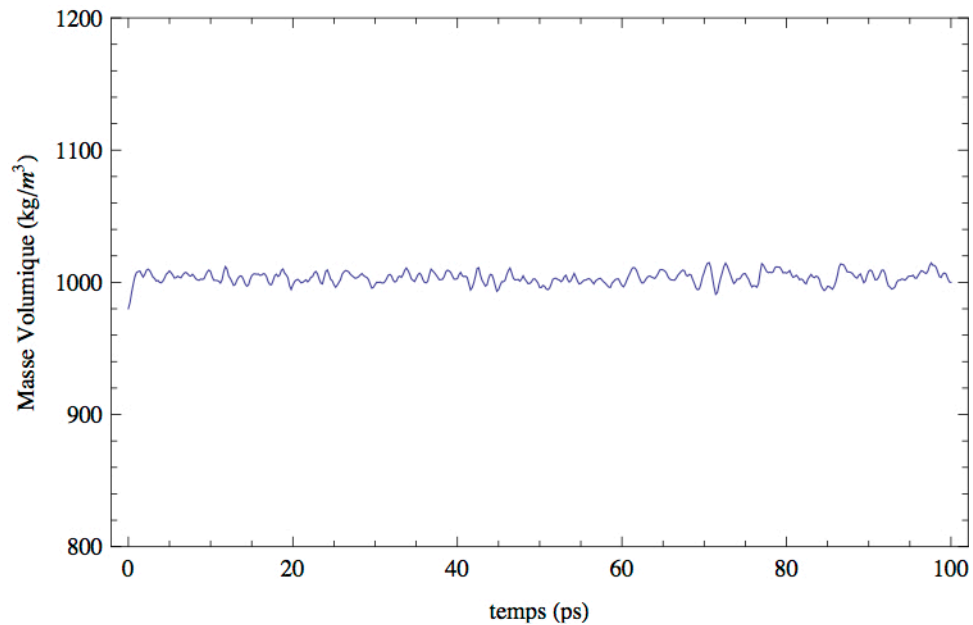


Figure 3.4 – Effet du barostat de Parinello-Rahman sur la masse volumique d'un système aqueux. La masse volumique est rapidement ajustée et conservée à  $1000 \text{ kg/m}^3$ .

#### 3.4.8. Phase d'équilibration

La phase d'équilibration de tous les systèmes se subdivise en deux étapes.

1. D'une part, un ensemble NVT a été obtenu en utilisant la méthode *V-Rescale*<sup>78</sup> implémenté dans GROMACS. Cette méthode est l'équivalent d'un thermostat de Berendsen avec l'ajout d'un terme stochastique permettant d'obtenir un ensemble canonique. La dynamique a été faite sur 100 ps à 300 K avec un incrément temporel de 0,002 ps.

2. Ensuite, l'ensemble NPT a été obtenu en utilisant le barostat Parrinello-Rahman<sup>77</sup>. La dynamique a été faite sur 100 ps à 300 K en conservant une pression de 1 bar avec un incrément temporel de 0,002 ps. Les constantes de compressibilités suivantes ont été utilisées :

Tableau IV – Facteurs de compressibilité utilisés au cours des dynamiques

Système	Solvant	Compressibilité (bar-1)
Trimère	eau	4,591 <sup>79</sup>
Trimère	THF	8,26 <sup>80</sup>
Rotaxane-parapluie	eau	4,591 <sup>79</sup>
Rotaxane-parapluie	dichlorométhane	10,3 <sup>79</sup>
Rotaxane-parapluie	eau/dichlorométhane	7,5

#### 3.4.9. Phase de production

La dernière étape consistait en la production. C'est à ce moment que l'information des dynamiques a été enregistrée. On suppose qu'à cette étape, l'énergie est conservée et le système représente un ensemble canonique. Dans le cadre des dynamiques moléculaires de ce chapitre, les étapes de production étaient toutes de 1000 ps. Les mêmes thermostats, barostats et incréments de temps ont été une fois de plus utilisés.

### 3.5. Simulation de l'encapsulation de pyrène par un trimère d'un dérivé d'acide cholique

Les acides choliques et leurs dérivés sont des molécules importantes dans les processus biologiques. Ils interviennent notamment dans la formation des sels biliaires, qui forment les micelles permettant de solubiliser les gras et le cholestérol dans le foie<sup>81</sup>. La structure particulière de l'acide cholique (Figure 3.5), avec des groupements hydroxyles hydrophiles

sur une face et des groupements méthyles hydrophobes sur l'autre, lui confère un comportement amphiphile. Cette caractéristique, couplée à la disponibilité commerciale du composé, en fait un précurseur intéressant pour la fabrication de nouveaux matériaux en chimie supramoléculaire<sup>82</sup>.

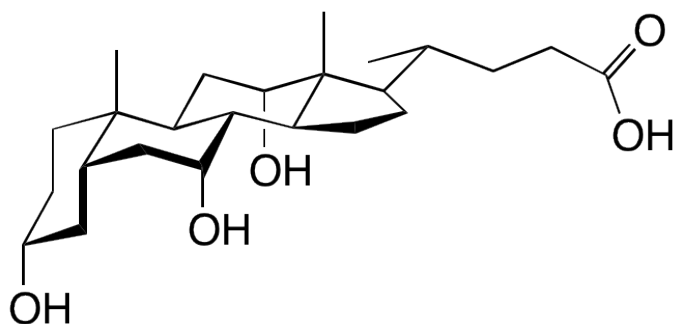


Figure 3.5 — Structure de l'acide cholique. On peut voir sur la face supérieure les groupements méthyles hydrophobes et sur la face opposée les groupements hydroxyles hydrophiles.

Dans cette section, nous allons voir comment la dynamique moléculaire a permis de simuler l'encapsulation de différents types de pyrène dans un trimère d'acides choliques. Cette étude fera la comparaison systématique avec les résultats expérimentaux publiés par Juntao Luo, Yilong Chen et Julian X. Zhu<sup>82c</sup>.

### 3.5.1. Sommaire des résultats expérimentaux disponibles

Les simulations ont été réalisées avec le trimère carbamide de l'acide cholique **3.1** (Figure 3.6). Il était question de voir les interactions entre le trimère et des espèces hydrophiles et hydrophobes par spectroscopie. Deux sondes ont donc été utilisées.



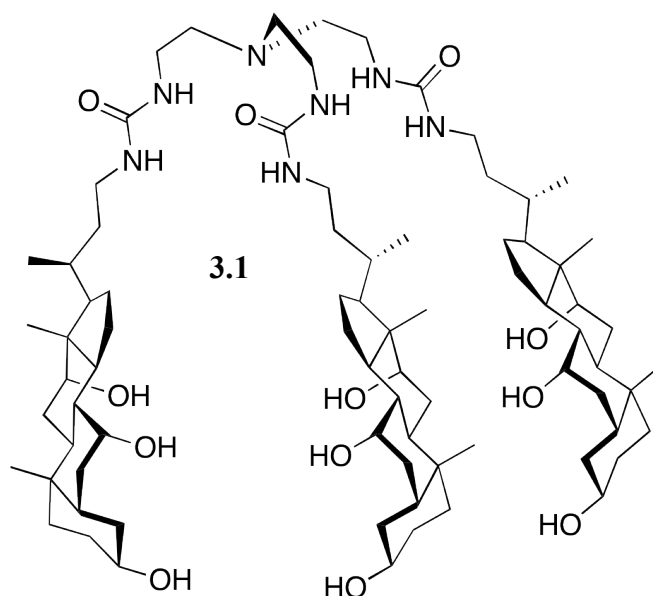


Figure 3.6 – Trimère carbamide de l'acide cholique.

L'absorbance du HPTS-OH **3.2** et la fluorescence du pyrène **3.3** (Figure 3.7) ont été observées pour montrer qu'en fonction de la polarité du solvant, les interactions entre le trimère sont davantage favorisées avec une sonde ou l'autre.

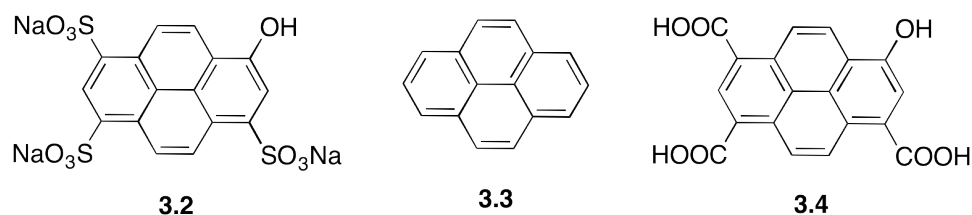


Figure 3.7 – Structure des sondes utilisées pour les analyses spectroscopiques.

La Figure 3.8 montre comment le signal des sondes change en présence du trimère en fonction de la polarité du solvant (PBS étant un tampon phosphate salin, la polarité du milieu augmente avec l'abscisse). Pour ce faire, l'absorbance du pyréne **3.2** et la fluorescence du composé **3.3** ont été observées. D'une part, on y voit qu'en absence du

trimère, l'absorption du pyrène **3.2** diminue en milieu polaire (courbe *a*) alors que le ratio de fluorescence du composé **3.3** diminue en milieu apolaire (courbe *c*). Toutefois, lorsque le trimère **3.1** est mis en solution, on voit un rehaussement soudain du signal d'absorption du composé **3.2** en milieu polaire (courbe *b*) de même qu'un changement brusque du signal de fluorescence du composé **3.3** en milieu apolaire (courbe *d*). Ce changement soudain confirme que le trimère change l'environnement local des sondes, supposant qu'il encapsule les sondes, limitant leur interaction avec le solvant.

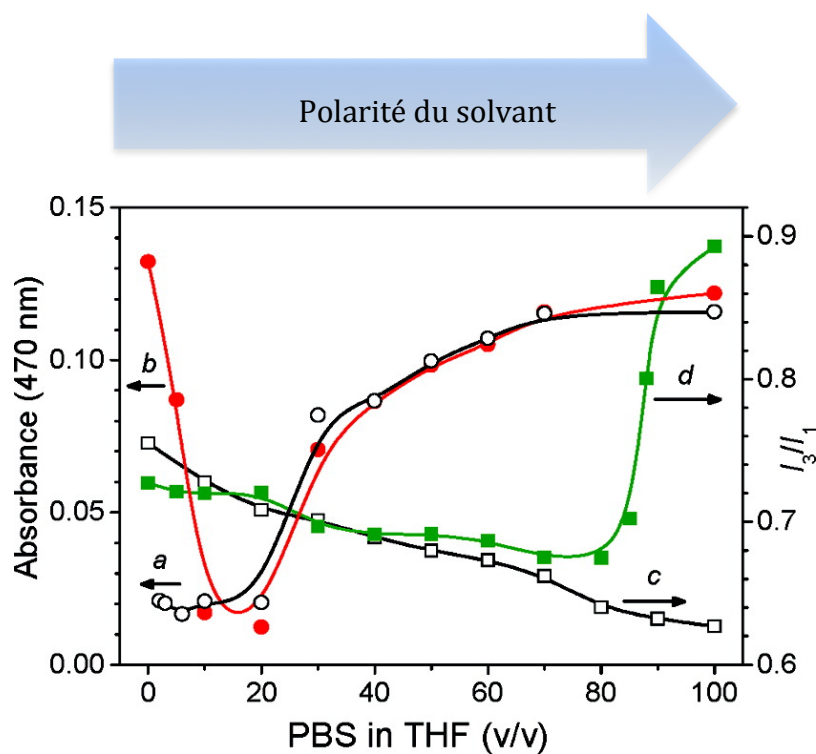


Figure 3.8 – Absorbance du composé **3.2** en absence et en présence du trimère **3.1** (courbe *a* et *b* respectivement) et fluorescence du composé **3.3** en absence et en présence du trimère **3.1** (courbes *c* et *d*)<sup>82c</sup>.

### 3.5.2. Résultats computationnels

Pour confirmer l'encapsulation des sondes par le trimère dans différents solvants, des dynamiques moléculaires ont été réalisées selon la méthodologie précédemment discutée.

Quatre systèmes ont été construits. D'une part, une molécule du trimère a été solvatée dans l'eau et dans le THF. Ensuite, une molécule de pyrène a été ajoutée dans chaque système et de la même façon, une molécule d'un dérivé polaire du pyrène (structure **3.4** de la Figure 3.7) a été ajoutée dans chaque système. Les spécifications des systèmes sont présentées dans le Tableau V.

Tableau V – Spécifications des systèmes du trimère

Nom du système	Sonde	Solvant	Molécules de solvant
PYR-H <sub>2</sub> O	<b>3.3</b>	H <sub>2</sub> O	7076
PYR-THF	<b>3.3</b>	THF	1416
HPT-H <sub>2</sub> O	<b>3.4</b>	H <sub>2</sub> O	7068
HPT-THF	<b>3.4</b>	THF	1415

L'analyse de la dynamique a été faite en calculant les centres géométriques du trimère et de la sonde et en soustrayant les deux valeurs pour chaque conformation. Dans le cas d'une sonde parfaitement encapsulée, on obtiendrait un  $\delta \rightarrow 0$  (les centres géométriques se confondent) et dans le cas d'une sonde libre loin du trimère, on obtiendrait une valeur  $\delta$  très élevée (les centres géométriques sont très éloignés). La Figure 3.9 montre les valeurs pour la sonde **3.3** dans l'eau et le THF.

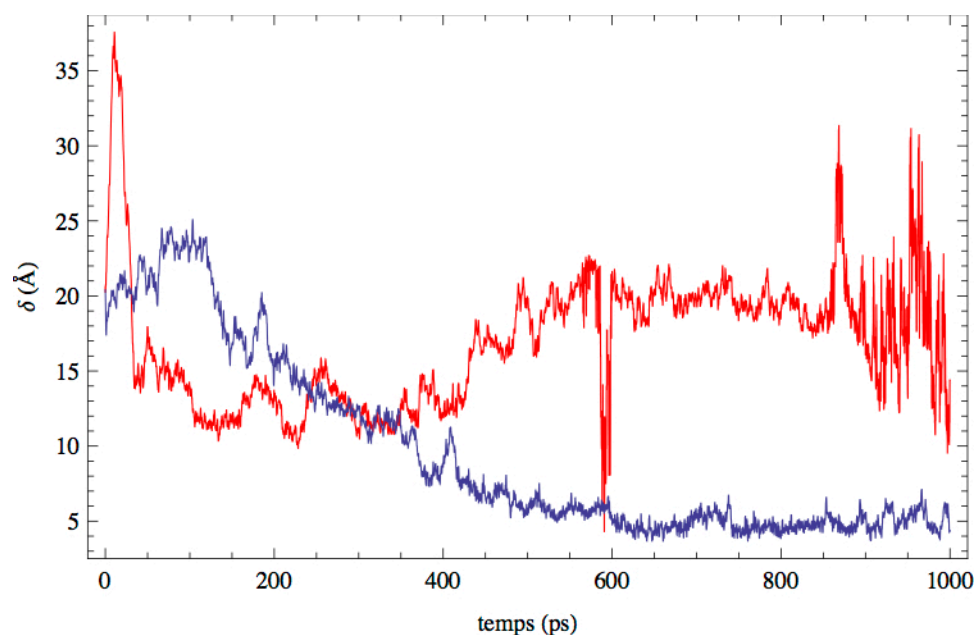


Figure 3.9 – Valeurs  $\delta$  entre la sonde **3.3** et le trimère **3.1** dans l'eau (en bleu) et dans le THF (en rouge).

D'une part, la Figure 3.9 montre une tendance nette vers un rapprochement du pyrène **3.3** et du trimère **3.1** en solution aqueuse dès 100 ps. Par ailleurs, nous pouvons remarquer que le pyrène et le trimère se rapprochent jusqu'à une distance  $\delta$  d'environ 5 Å et demeurent à cette distance pour le reste de la simulation. La Figure 3.10 montre la conformation finale et confirme l'encapsulation de la sonde **3.3** par le trimère **3.1** dans l'eau :

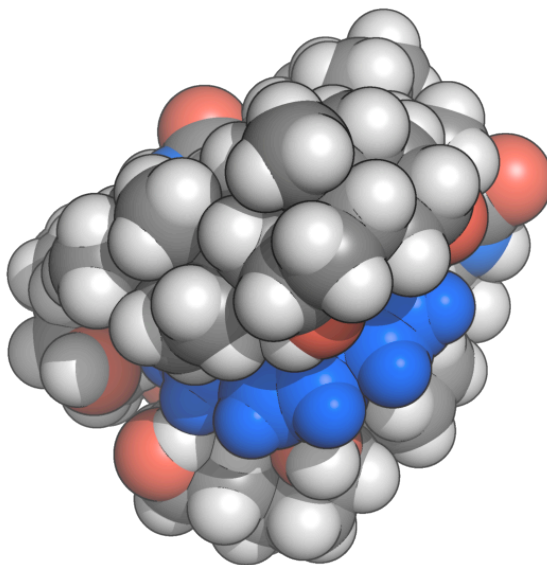


Figure 3.10 – Conformation finale du pyrène **3.3** et du trimère **3.1** dans l'eau. Le pyrène **3.3** est illustré en bleu et le trimère **3.1** est illustré selon les couleurs usuelles des représentations atomiques.

Nous pouvons par ailleurs constater dans la Figure 3.9 que le pyrène **3.3** a tendance à s'éloigner du trimère **3.1** dès 400 ps, restant librement dans la solution. En effet, la distance moyenne entre les deux composés est d'environ 20 Å, indiquant que les interactions entre le pyrène **3.3** et le trimère sont limitées. L'illustration de la conformation finale (Figure 3.11) montre comment les interactions entre le trimère **3.1** et le pyrène **3.3** sont limitées. Le même type d'expérience a été réalisé avec les systèmes HPT-H<sub>2</sub>O et HPT-THF. La Figure 3.12 compare les valeurs  $\delta$  obtenues pour les deux systèmes.

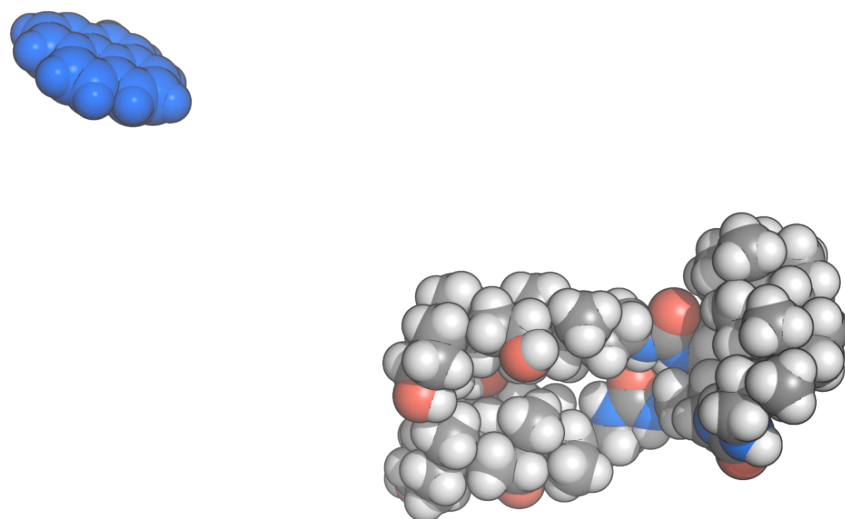


Figure 3.11 – Conformation finale du pyrène **3.3** et du trimère **3.1** dans le THF. Le pyrène **3.3** est illustré en bleu alors que le trimère **3.1** est illustré selon les couleurs usuelles des représentations atomiques.

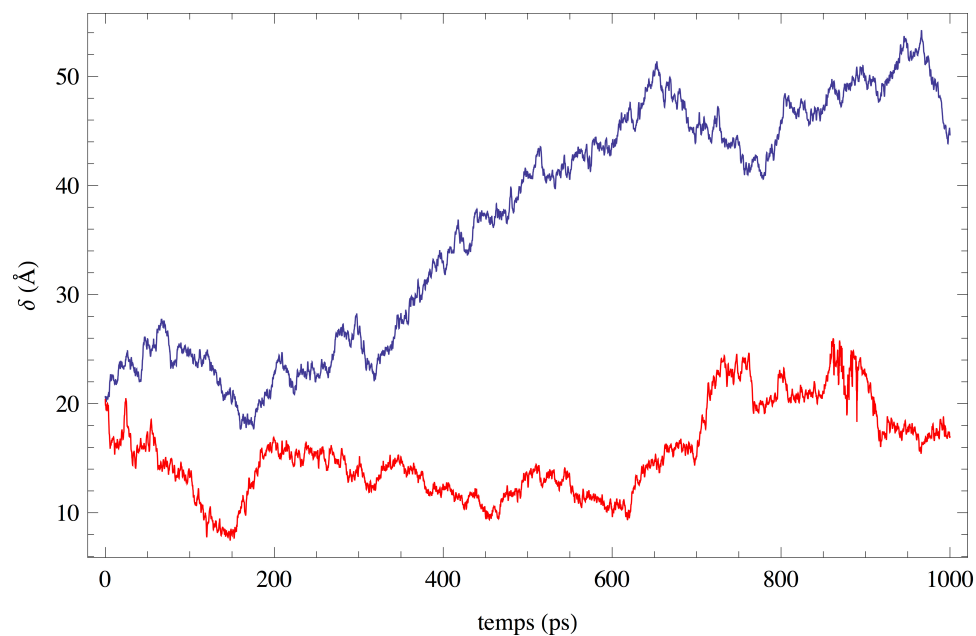


Figure 3.12 – Valeur  $\delta$  entre la sonde **3.4** et le trimère **3.1** dans l'eau (en bleu) et dans le THF (en rouge).

D'une manière similaire aux systèmes PYR-H<sub>2</sub>O et PYR-THF, on peut voir pour ces deux nouveaux systèmes des différences notoires entre les distances  $\delta$  entre la sonde **3.4** et le trimère. D'une part, nous pouvons constater qu'en milieu aqueux, la sonde et le trimère s'éloignent et restent bien séparés (la valeur  $\delta$  moyenne est d'environ 48 Å). La Figure 3.13 confirme que la sonde et le trimère n'interagissent pas dans l'eau.

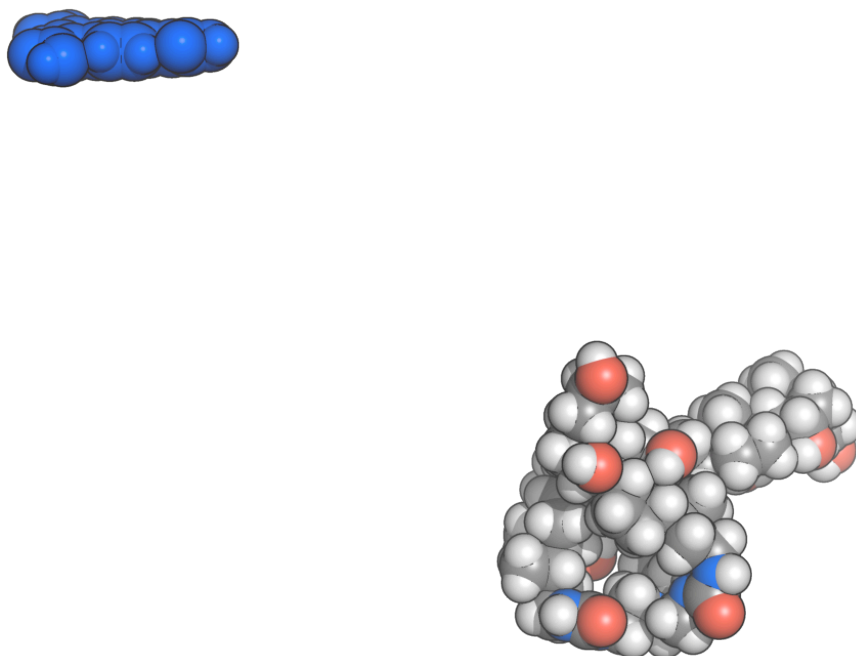


Figure 3.13 – Conformation finale du pyrène **3.4** et du trimère **3.1** dans l'eau. Le pyrène **3.4** est illustré en bleu alors que le trimère **3.1** est illustré selon les couleurs usuelles des représentations atomiques.

Le système PYR-THF (illustré en rouge dans la Figure 3.12) montre cette fois-ci une plus grande interaction entre la sonde et le trimère. On notera que la valeur  $\delta$  reste quand même plus grande que celle du système PYR-H<sub>2</sub>O. En effet, du fait du volume moléculaire plus important de la sonde **3.4** par rapport à la sonde **3.3**, l'encapsulation totale par le trimère est plus difficile, et les centres géométriques des composés sont plus éloignés. La Figure 3.14 montre que l'encapsulation est partielle cette fois-ci.

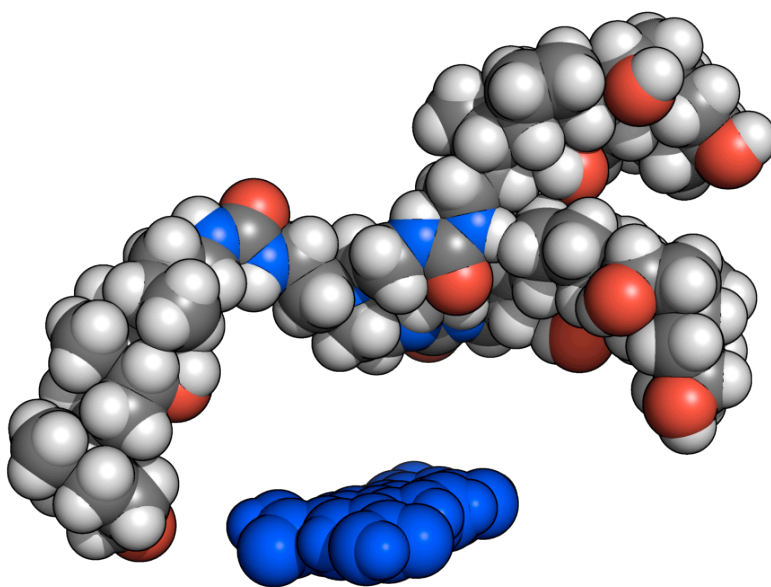


Figure 3.14 – Conformation finale du pyrène **3.4** et du trimère **3.1** dans le THF. Le pyrène **3.4** est illustré en bleu alors que le trimère **3.1** est illustré selon les couleurs usuelles des représentations atomiques.

### 3.5.3. Conclusions

En somme, la méthode de dynamique moléculaire développée dans le contexte de cette étude simule de façon qualitative les processus supramoléculaires d'agrégation dans différents solvants. Il est désormais clair que la structure amphiphile du trimère lui permet un réarrangement dans différents solvants, favorisant par la même occasion certains processus d'encapsulation de molécules compatibles.

Ces résultats nous ont encouragés à poursuivre nos analyses sur des systèmes plus complexes. Nous verrons dans la section suivante comment il a été possible, par la même méthodologie, de construire une séparation de phase entre deux solvants et de simuler le passage d'un rotaxane d'un environnement à un autre.



### 3.6. Simulation des interactions d'un rotaxane-parapluie à l'interface d'une séparation de phase entre l'eau et le dichlorométhane

La perméabilité des membranes biologiques est un obstacle important à l'élaboration de nouveaux médicaments efficaces<sup>83</sup>. Dans beaucoup de cas, le composé sera ou trop polaire et donc bloqué par la membrane, ou trop peu polaire et peu soluble dans l'eau. Parmi les méthodes suggérées pour contourner ce problème, l'utilisation d'une molécule amphiphile transporteuse a été proposée<sup>84</sup>. La molécule proposée, un rotaxane-parapluie est constituée d'une roue entourant un axe. L'assemblage est mécaniquement tenu en place par des groupements bloquant sur l'axe qui empêchent, par encombrement stérique, la roue de se détacher (Figure 3.15).

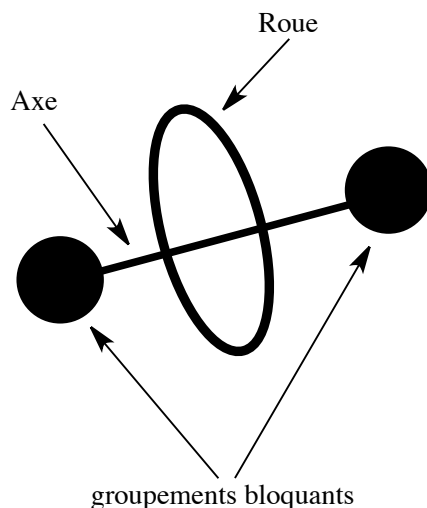


Figure 3.15 – Schéma d'un rotaxane

Nous verrons dans cette section comment la dynamique moléculaire a pu partiellement reproduire certains résultats expérimentaux et donner de nouvelles indications sur le processus de transport d'un rotaxane dans une membrane phospholipidique.

### 3.6.1. Sommaire des résultats expérimentaux disponibles

Les molécules **3.5** et **3.6** ont été synthétisées (Figure 3.16) et assemblées par Christine Chhun pour donner le rotaxane **3.5** + **3.6**. De même, les propriétés de transport membranaire de ce système ont été démontrées<sup>84</sup>.

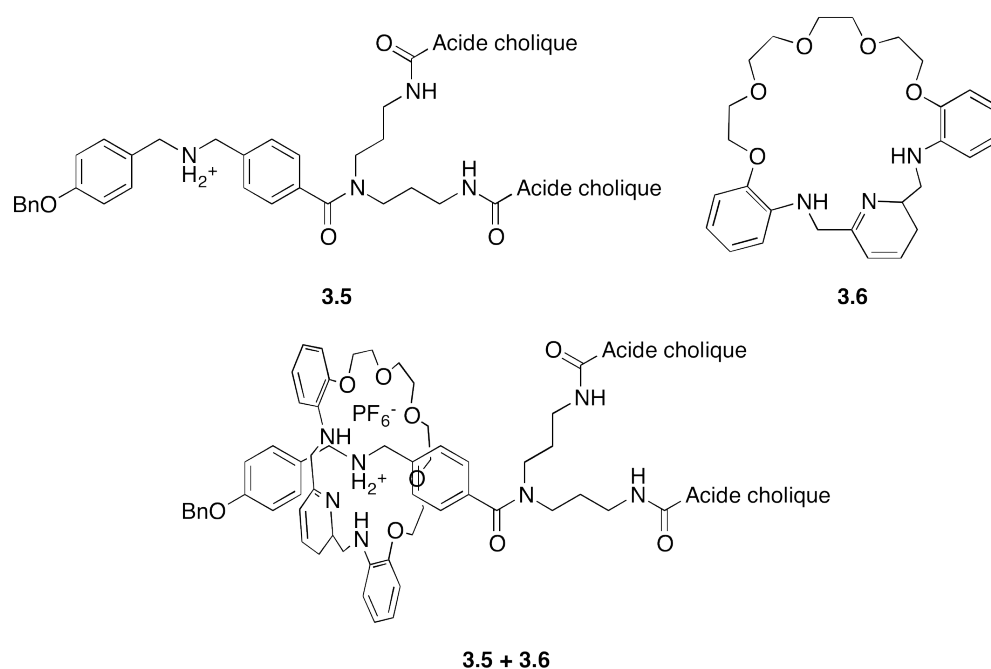


Figure 3.16 – Illustration de l'axe **3.5**, de la roue **3.6** et de l'assemblage supramoléculaire du rotaxane-parapluie **3.5** + **3.6**.

Des tests de transport dans des liposomes d' $\alpha$ -phosphatidylcholine contenant une solution de leucigénine ont permis de montrer l'efficacité des rotaxanes-parapluie comme transporteurs d'anions chlorure<sup>85</sup>. Afin d'approfondir notre compréhension du phénomène, nous avons réalisé des dynamiques moléculaires du rotaxane-parapluie à l'interface dans un système biphasique.

### 3.6.2. Résultats computationnels

La simulation de la séparation de phase a été réalisée en plusieurs étapes. D'une part, afin d'avoir une idée de la conformation privilégiée par le rotaxane dans l'un ou l'autre des solvants, deux dynamiques ont été faites : une première dans l'eau et une deuxième dans le dichlorométhane. Pour des raisons de faisabilité des dynamiques, le dichlorométhane, par son comportement lipophile, a été utilisé comme modèle pour simuler un environnement phospholipidique. De plus, le contre-ion du rotaxane-parapluie n'a pas été simulé. La distance entre les bras et le cœur polaire de **3.5** + **3.6** a été observée au cours des deux dynamiques afin de suivre le changement de conformation du rotaxane (Figure 3.17). La Figure 3.17 montre clairement que la distance entre les bras et le cœur polaire est plus petite en milieu lipophile, et donc que dans ce milieu, le cœur est protégé par les bras amphiphiles. L'observation des structures majoritaires (Figure 3.18 et Figure 3.19) confirme ce phénomène.

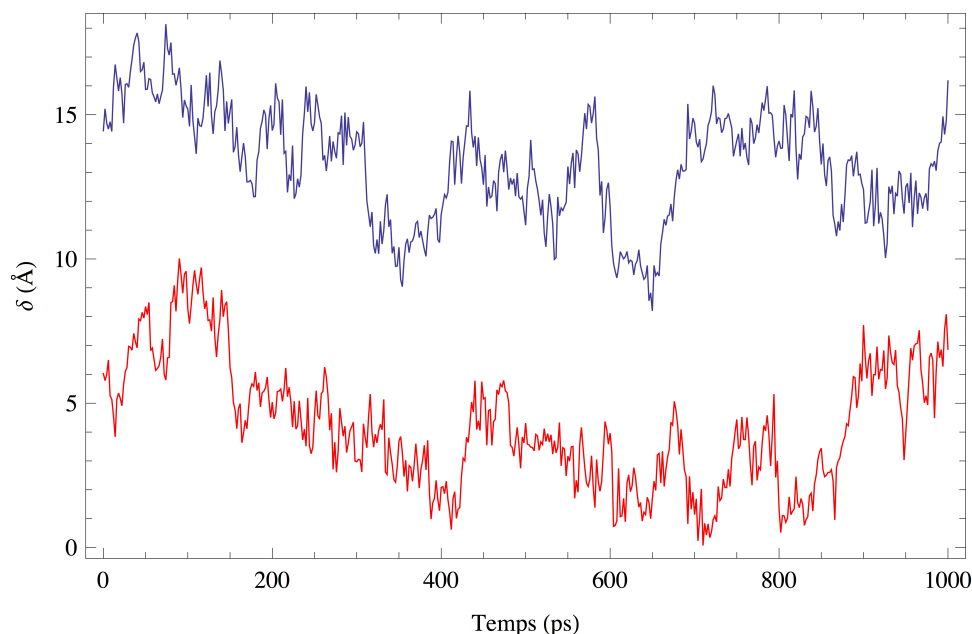


Figure 3.17 – Valeurs  $\delta$  entre les bras et le cœur polaire du rotaxane **3.5** + **3.6**. L'évolution du  $\delta$  dans le chloroforme est illustrée en rouge et celle dans l'eau est illustrée en bleu.

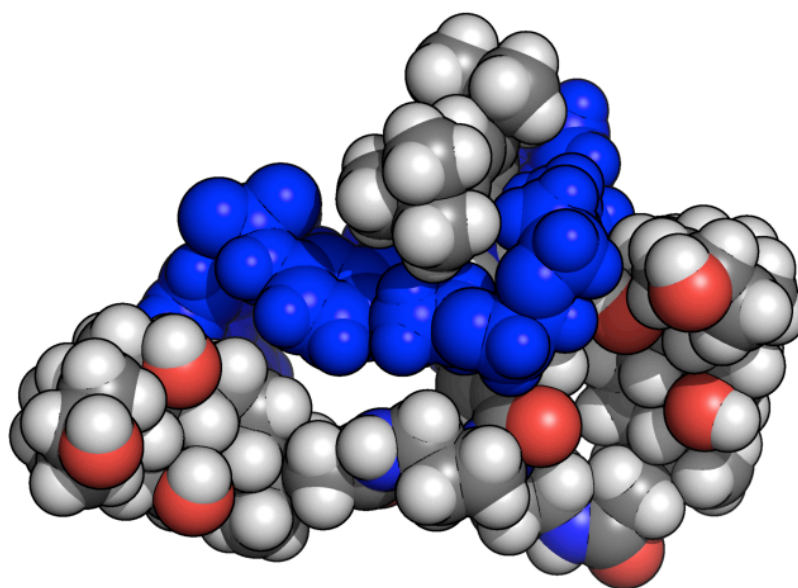


Figure 3.18 – Conformation du rotaxane-parapluie **3.5** + **3.6** dans le dichlorométhane. La roue, illustrée en bleue, est recouverte par les acides choliques.

Lorsque le rotaxane-parapluie **3.5** + **3.6** est placé dans le dichlorométhane, nous pouvons constater que la roue est protégée par la partie hydrophile des acides choliques. Ce comportement est inversé lorsque le rotaxane est placé dans l'eau (Figure 3.19).

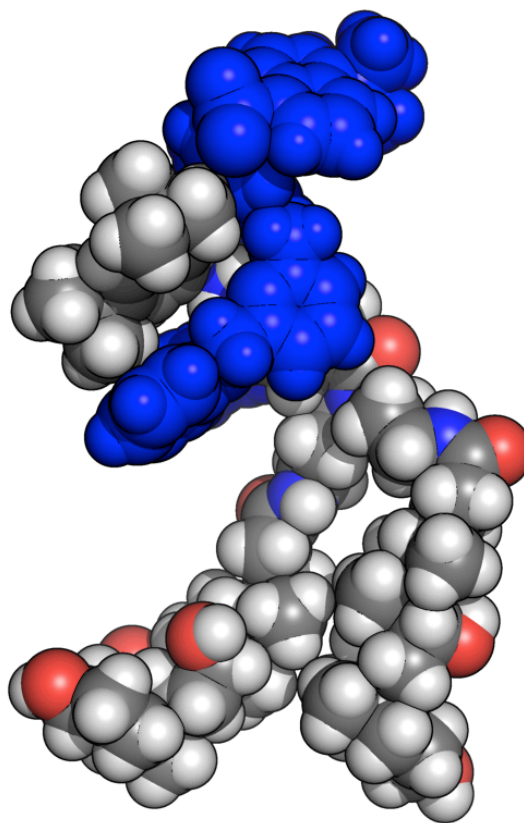


Figure 3.19 – Conformation du rotaxane-parapluie **3.5 + 3.6** dans l'eau. La roue, illustrée en bleue, est complètement exposée au solvant et la partie hydrophobe des acides choliques est regroupée pour limiter son interaction avec le solvant.

On peut voir que la roue est bien exposée au solvant. Par ailleurs, les faces hydrophobes des acides choliques sont regroupées alors que les faces hydrophiles sont exposées au solvant. Pour tenter de mieux comprendre le comportement du rotaxane à l'interface d'une bicouche, un système biphasique composé de dichlorométhane et d'eau a été construit. Ensuite, deux dynamiques ont été faites, la première en plaçant le rotaxane **3.5 + 3.6** dans la partie aqueuse et la seconde en le plaçant dans le solvant organique (Figure 3.20). Le Tableau VI donne la composition des systèmes.

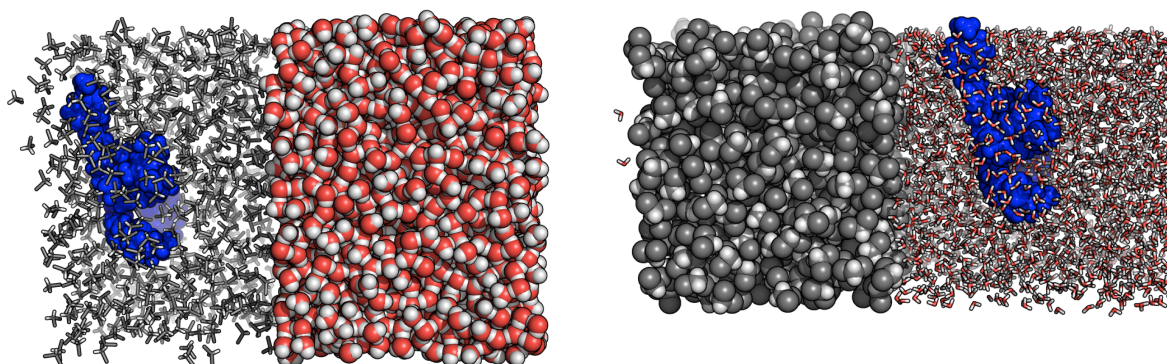


Figure 3.20 – Conformations initiales des simulations de séparation de phase. Le rotaxane **3.5** + **3.6** est illustré en bleu. À gauche, on voit le rotaxane dans le dichlorométhane et à droite, on voit le rotaxane dans l'eau. Pour plus de clarté, la représentation en tige a été choisie pour les solvants contenant le rotaxane.

Tableau VI – Spécification des systèmes biphasiques

Nom du système	Milieu contenant 4+5	Nb molécules H <sub>2</sub> O	Nb molécules CH <sub>2</sub> Cl <sub>2</sub>
<b>3.5</b> + <b>3.6</b> –H <sub>2</sub> O	H <sub>2</sub> O	2386	503
<b>3.5</b> + <b>3.6</b> –CH <sub>2</sub> Cl <sub>2</sub>	CH <sub>2</sub> Cl <sub>2</sub>	3462	681

Afin d'observer le déplacement du rotaxane, la position moyenne de **3.5** + **3.6** sur l'axe z a été enregistrée et portée en graphique. L'axe z est coupé par les deux phases dans la Figure 3.20. La Figure 3.21 et la Figure 3.22 montrent l'évolution de la position du rotaxane **3.5** + **3.6** initialement placé dans le dichlorométhane et dans l'eau respectivement.

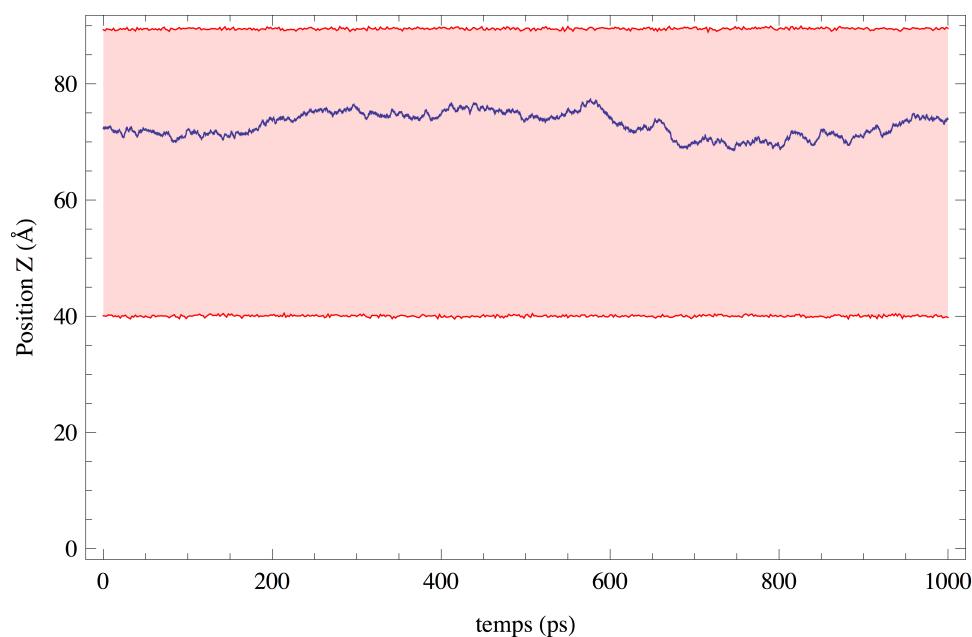


Figure 3.21 – Trajectoire du rotaxane **3.5 + 3.6** initialement placé dans la partie lipophile du système biphasique. L'axe des ordonnées représente la composante z des coordonnées atomiques des molécules. Le tracé en bleu représente la trajectoire du rotaxane. Les tracés rouges cernent les coordonnées minimales et maximales des molécules de dichlorométhane.

La Figure 3.21 montre clairement que le rotaxane reste à peu près à la même position durant toute la dynamique. Ce comportement n'est toutefois pas observé lorsque le rotaxane **3.5 + 3.6** est initialement placé en milieu aqueux (Figure 3.22).

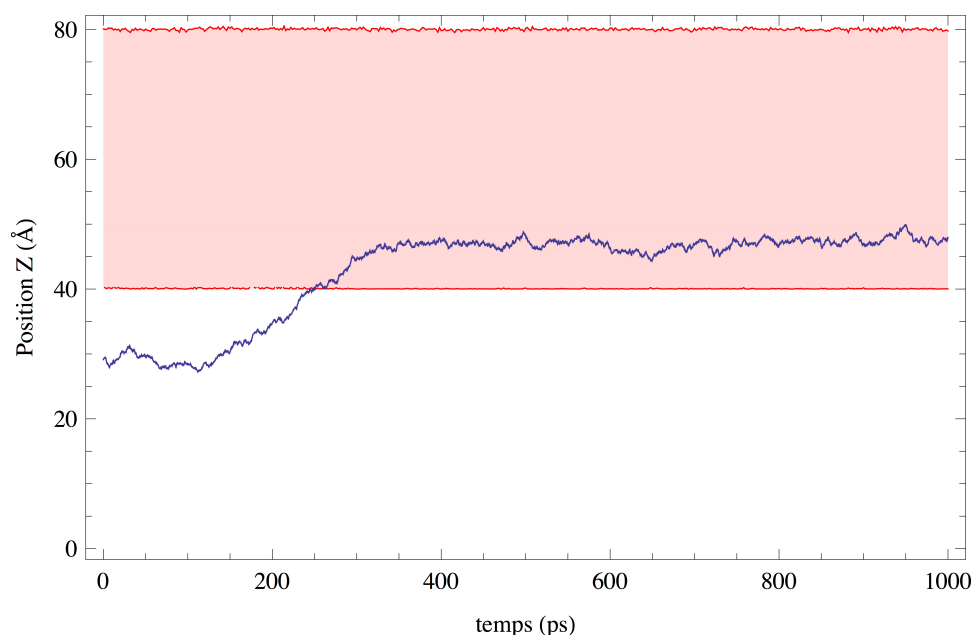


Figure 3.22 – Trajectoire du rotaxane **3.5 + 3.6** initialement placé dans la partie hydrophile du système biphasique. L'axe des ordonnées représente la composante z des coordonnées atomiques des molécules. Le tracé en bleu représente la trajectoire du rotaxane. Les tracés rouges cernent les coordonnées minimales et maximales des molécules de dichlorométhane.

En effet, la Figure 3.22 montre un transfert de phase du rotaxane dès 220 ps. Cette différence suggère que le caractère amphiphile du rotaxane **3.5 + 3.6** permet le transport d'un milieu hydrophile vers un milieu lipophile et cette dynamique a permis de suggérer un mécanisme probable pour le transport. Quatre étapes sont proposées pour expliquer le passage du rotaxane d'un milieu aqueux vers un milieu lipophile.

### 1. Approche d'un bras vers le milieu lipophile

Initialement, au moins un des bras se rapproche de l'interface entre les milieux (Figure 3.23). La figure montre un des bras (à droite dans la figure) qui se place à l'interface.



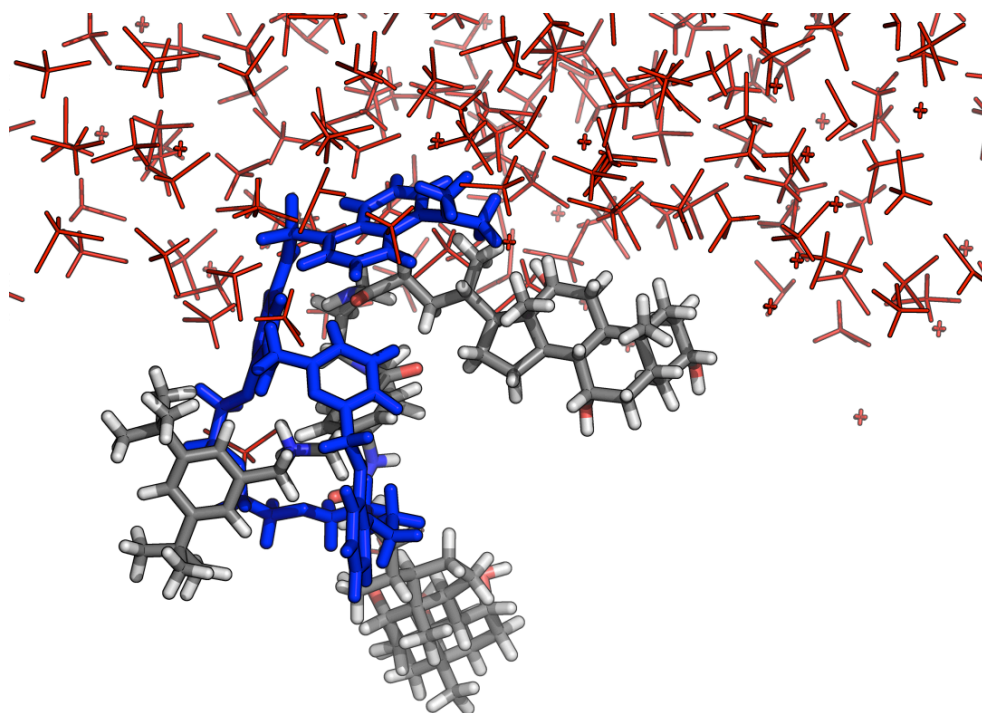


Figure 3.23 – Rapprochement d'un des bras du rotaxane **3.5** + **3.6** à l'interface. Pour clarifier l'illustration, l'eau a été masquée. Le dichlorométhane est illustré en rouge, la roue est illustrée en bleu et l'axe est illustré selon les couleurs usuelles des représentations atomiques. Le bras à droite s'est rapproché de l'interface. Temps de la dynamique :  $t = 220$  ps. Les croix font référence à des représentations incomplètes de molécules qui sortent de la boîte.

## 2. Rapprochement du cœur et des groupements polaires du bras à l'interface

Par la suite, le bras placé à l'interface et le cœur polaire se rapprochent l'un de l'autre (Figure 3.24).

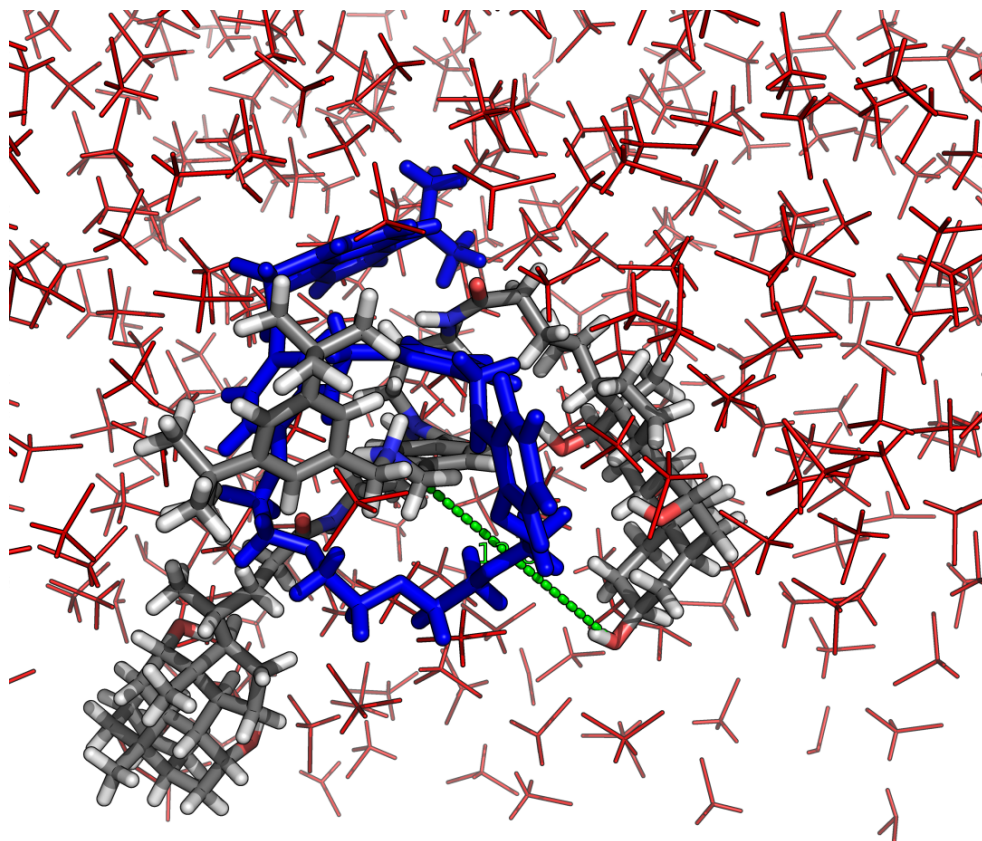


Figure 3.24 – Illustration du cœur polaire et de la portion hydrophile de l'acide cholique à l'interface. Pour clarifier l'illustration, l'eau a été masquée. Le dichlorométhane est illustré en rouge, la roue est illustrée en bleu et l'axe est illustré selon les couleurs usuelles des représentations atomiques. On voit que le cœur polaire est près de l'acide cholique à droite. Temps de la dynamique :  $t = 300$  ps.

La figure montre le bras de droite près du cœur polaire. Sur un intervalle de temps d'environ 100 ps, la distance  $\delta$  entre le cœur polaire de l'axe et le groupement hydroxyle à l'extrémité de l'acide cholique diminue de près de 4 Å, ce qui montre que la roue se rapproche du bras à l'interface (Figure 3.25).

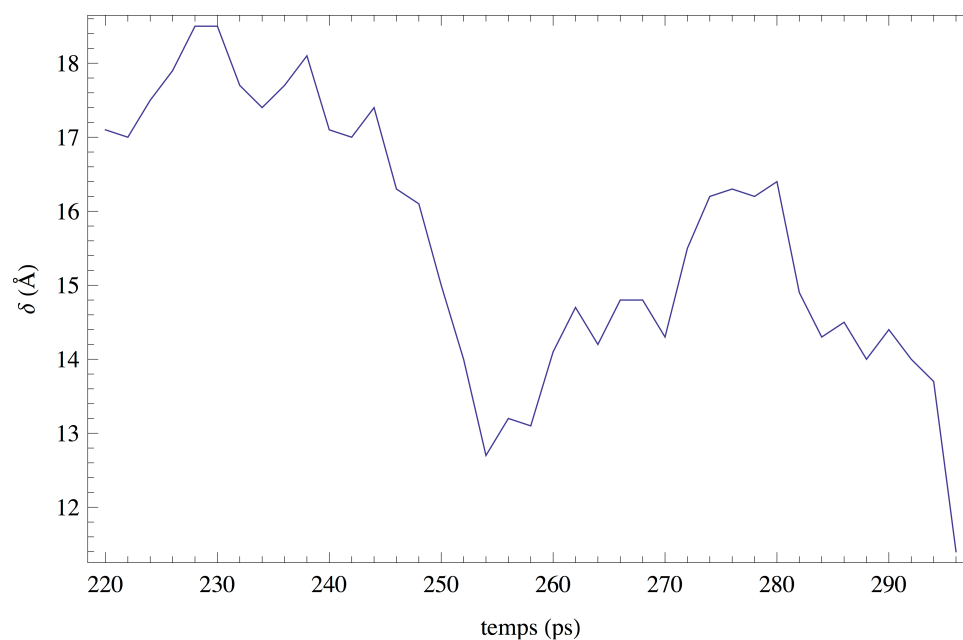


Figure 3.25 – Évolution de la distance entre le bras et le cœur polaire du rotaxane **3.5 + 3.6**. La valeur  $\delta$  représente la distance entre le cœur polaire et l'oxygène du groupement hydroxyle à l'extrémité de l'acide cholique à l'interface.

### 3. Rapprochement du deuxième bras à l'interface

Par la suite, on peut voir le deuxième bras, qui jusqu'à maintenant se déplaçait librement dans l'eau, approcher et se placer à l'interface.

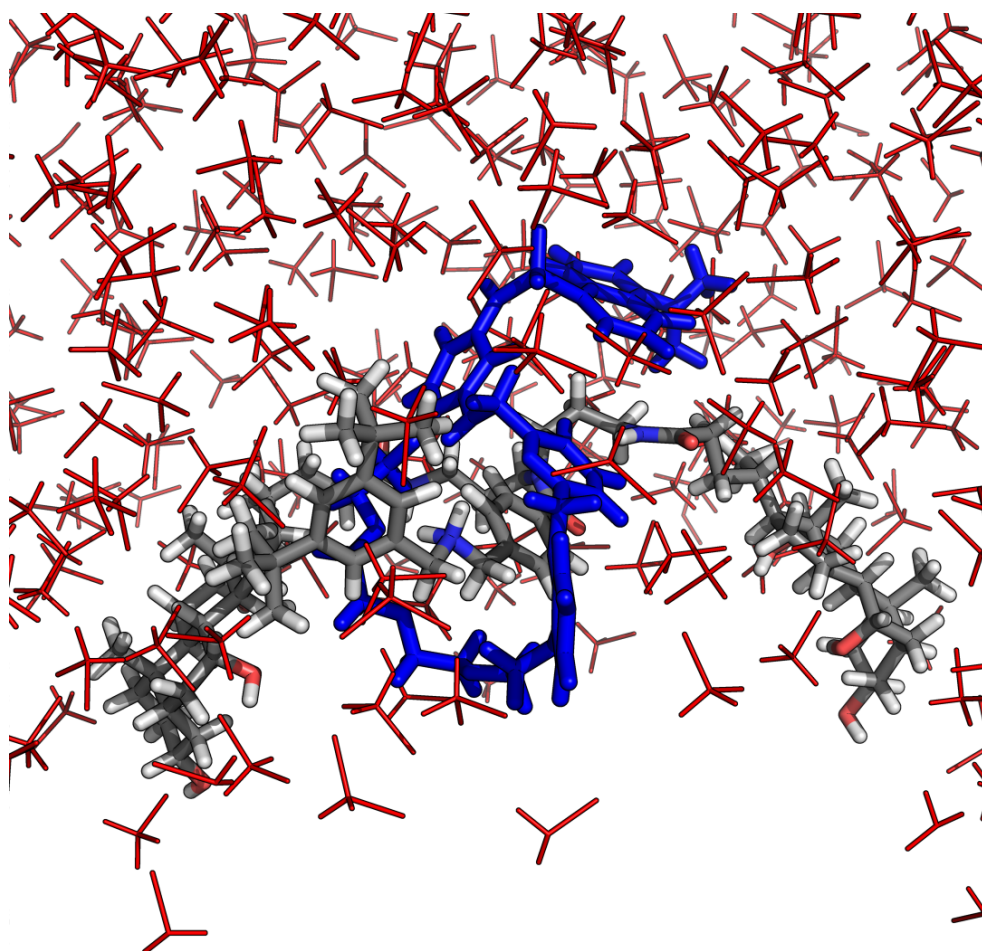


Figure 3.26 – Rapprochement du second bras à l'interface. Pour clarifier l'illustration, l'eau a été masquée. Le dichlorométhane est illustré en rouge, la roue est illustrée en bleu et l'axe est illustré selon les couleurs usuelles des représentations atomiques. On voit que le second bras, visible à gauche dans l'image est désormais placé à l'interface des deux solvants. Temps de la dynamique :  $t = 400$  ps.

#### 4. Transition de phase du cœur polaire

Lorsque les bras sont placés à l'interface, le cœur polaire est en mesure de passer dans la phase lipophile.

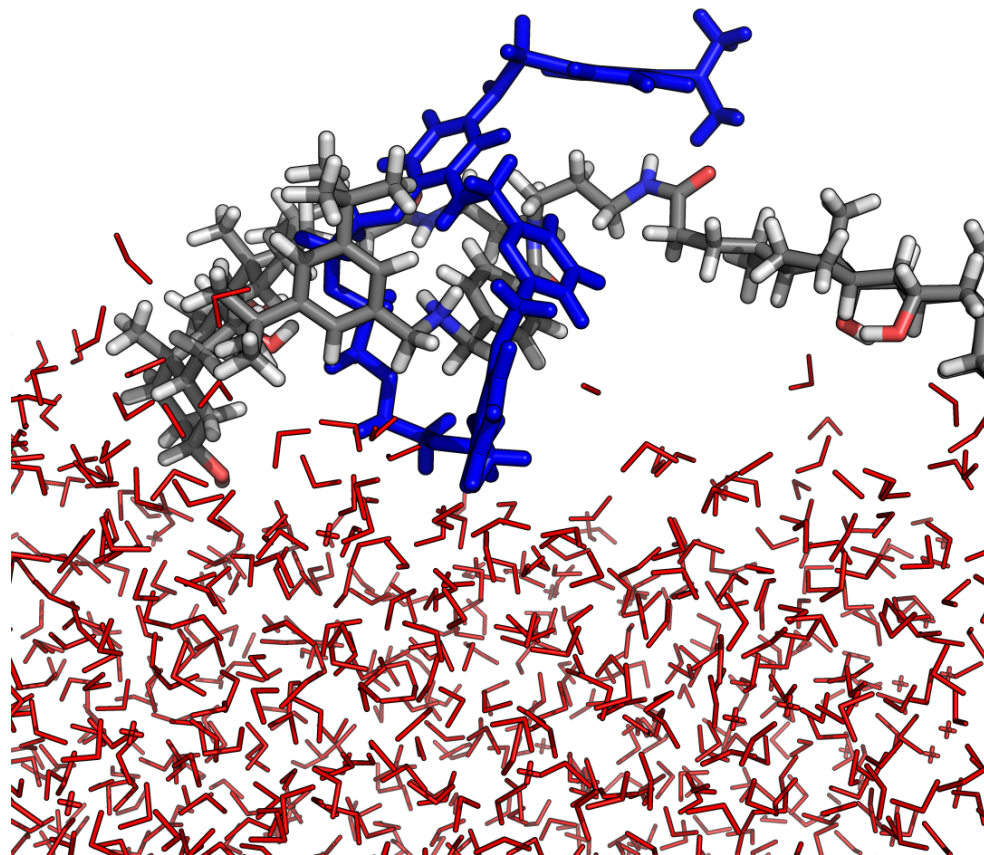


Figure 3.27 – Passage du cœur dans la phase lipophile. Pour clarifier l'illustration, le dichlorométhane a été masqué. L'eau est illustrée en rouge, la roue est illustrée en bleu et l'axe est illustré selon les couleurs usuelles des représentations atomiques. On voit que la roue est majoritairement solvatée dans le dichlorométhane. Temps de la dynamique :  $t = 800$  ps.

Les différentes simulations réalisées n'ont pas permis de constater qu'il y avait une encapsulation subséquente de la roue par les bras permettant le passage rapide du rotaxane à travers le milieu lipophile. Cette constatation suggère donc que le transport du rotaxane pourrait survenir, tel que proposé<sup>84</sup> par un mécanisme de dimérisation du rotaxane.

### 3.6.3. Limites et conclusions

En somme, le rotaxane parapluie **3.5 + 3.6** a été simulé dans divers environnements : l'eau, le dichlorométhane et un mélange des deux solvants. L'utilisation du dichlorométhane avait pour objectif de générer un modèle simplifié d'une membrane lipidique. Les simulations ont montré qu'en milieu aqueux, les bras du rotaxane sont plus éloignés du cœur qu'en milieu lipophile. Les conformations finales montrent d'ailleurs clairement comment d'un solvant à l'autre, la roue se fait enrober ou non par les bras d'acides choliques. Ces résultats nous ont encouragés à amorcer une simulation dans un système contenant un mélange biphasique des deux solvants. À partir d'une position initiale du rotaxane dans l'un ou l'autre des solvants, nous avons montré que le rotaxane, initialement placé dans l'eau, migrerait assez facilement dans le dichlorométhane malgré la présence d'un groupement central assez polaire. Un mécanisme de transport a été suggéré.

Bien que les résultats obtenus concordent assez bien avec les résultats expérimentaux disponibles présentés dans la section 3.6.1, le modèle développé ne répond pas à toutes les subtilités d'une bicouche lipidique. En effet, il est évident que le dichlorométhane est géométriquement et électroniquement très différent d'un phospholipide. Par ailleurs, contrairement aux systèmes simulés, les bicouches lipidiques possèdent un gradient d'hydrophobie de la surface vers le cœur. Les systèmes simulés proposent quant à eux un changement beaucoup plus brutal de la polarité du milieu. Nous pensons toutefois qu'en dépit de ces différences, les simulations réalisées permettent d'éclaircir le processus de transport du rotaxane **3.5 + 3.6**, du moins en ce qui concerne le changement de la phase aqueuse vers la phase lipophile de cette molécule.

### 3.7. Conclusions générales

De manière générale, nous pouvons conclure que les modèles mathématiques choisis ont permis de simuler de façon adéquate l'équilibre fragile entre les diverses interactions des

systèmes supramoléculaires que nous avons testés. D'une part, le choix de la dynamique moléculaire sur un intervalle de temps de 1000 ps nous a permis d'observer les phénomènes d'encapsulation de différents pyrènes dans un trimère en fonction de la polarité du solvant. De plus, nous avons pu observer les comportements découlant d'un changement de phase d'un rotaxane-parapluie. Étant donné que ces phénomènes ont pris forme sur une échelle de temps relativement élevée, la probabilité de les observer avec de simples méthodes d'optimisation de géométrie aurait été pratiquement nulle.

De plus, les conformations obtenues dans divers solvants montrent que les paramètres du champ de force GAFF sont adéquats pour simuler les systèmes organiques qui nous intéressent, à tous le moins de manière qualitative. Par ailleurs, la stabilisation de l'énergie potentielle et de la masse volumique montre l'efficacité des méthodes de solvation en conditions périodiques et du traitement des interactions non liantes sur des distances. Les thermostats et barostats utilisés ont permis dans tous les cas de stabiliser la température et la pression des systèmes afin d'obtenir l'environnement NPT souhaité. Finalement, malgré la taille importante des systèmes modélisés, les simplifications réalisées ont permis de compléter tous les calculs dans des délais raisonnables avec les infrastructures informatiques disponibles.

## **Chapitre 4: Utilisation de l'arrimage moléculaire comme stratégie de développement d'un nouvel inhibiteur enzymatique**

La dynamique moléculaire est un outil puissant qui nous permet de comprendre et même de prédire des phénomènes qui se produisent dans l'espace moléculaire et sur de petites échelles de temps. De plus, l'utilisation des champs de force et des équations de la mécanique classique constitue une solution élégante aux difficultés mathématiques imposées par les modèles quantiques décrits plus tôt.

Lorsqu'il est question d'étudier de très gros systèmes (plus de  $10^5$  atomes) ou des phénomènes évoluant sur une échelle de temps considérable (millisecondes), les méthodologies utilisées au chapitre précédent ne sont plus réalisables avec des infrastructures informatiques modestes; nous devons dès lors chercher à trouver de nouvelles approximations à faire pour calculer les phénomènes qui nous intéressent.

Dans ce chapitre, nous verrons comment il est possible de simuler de très grands systèmes par l'utilisation d'un algorithme génétique. Nous observerons les interactions entre une protéine et plusieurs molécules pour tenter de comprendre les facteurs moléculaires qui permettent de discriminer les bons des mauvais candidats.

Comme nous l'avons vu dans les chapitres précédents, la stratégie principale mise en place pour obtenir des conformations représentatives des observations expérimentales consiste à minimiser l'énergie des systèmes; c'est-à-dire à parcourir la surface de l'énergie potentielle pour tous les degrés de liberté du système en espérant trouver rapidement un creux de vallée correspondant à un minimum global (Figure 4.1).



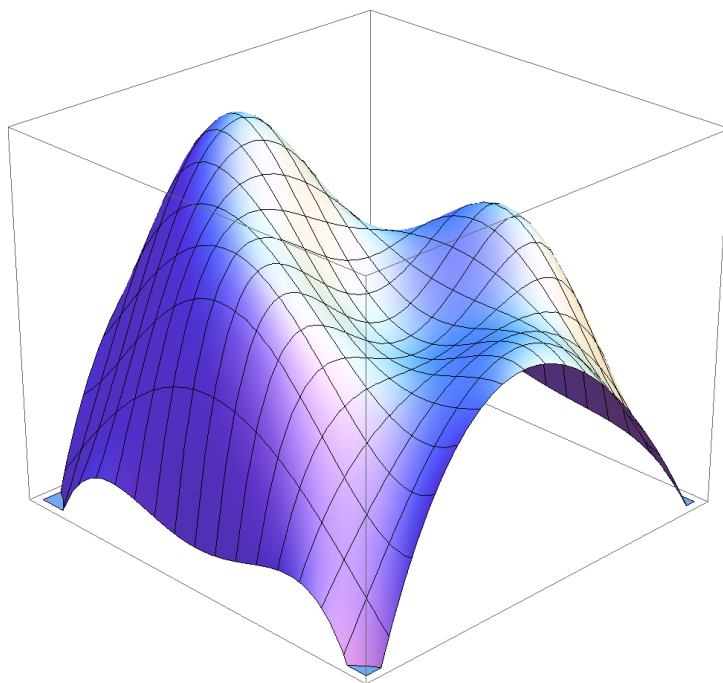


Figure 4.1 – Surface de potentiel pour deux degrés de liberté. L'axe des z représente l'énergie potentielle du système. L'axe des x et des y représente un degré de liberté. Les coins de la figure représentent des minima globaux, alors que le puits dans le centre de la surface représente un minimum local.

Cependant, à mesure que le nombre de degrés de liberté augmente, l'optimisation devient de plus en plus longue et fastidieuse. Les algorithmes de parcours systématique de la surface de potentiel sont donc remplacés par des algorithmes génétiques.

#### **4.1. Algorithmes génétiques et arrimage moléculaire**

Les algorithmes génétiques font partie de la classe des algorithmes évolutionnistes<sup>86</sup>. Comme toutes les métaheuristiques, ils ne garantissent pas d'obtenir la solution optimale au problème, mais peuvent résoudre les problèmes d'optimisation difficiles, voire impossibles à résoudre par d'autres méthodes<sup>87</sup>.

Proposés dès 1975 par John H. Holland<sup>88</sup>, les algorithmes génétiques appliqués à la modélisation moléculaire sont en quelque sorte l'extrapolation la plus poussée du premier principe évoqué au Chapitre 3 : soit que le système tend vers un état thermodynamiquement favorable. Le concept derrière les algorithmes génétiques se base sur l'analogie de l'évolution des espèces par la sélection naturelle.

Concrètement, les algorithmes génétiques fonctionnent de manière itérative selon le schéma suivant :

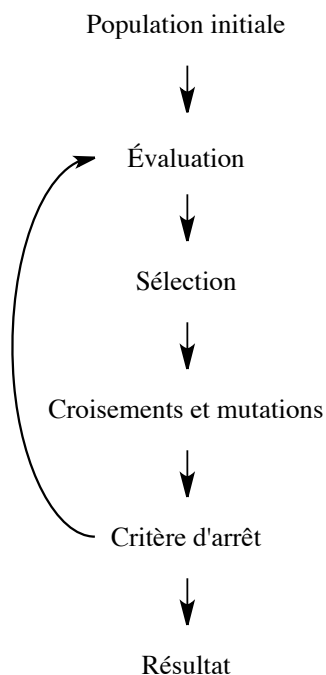


Figure 4.2 – Schéma itératif typique d'un algorithme génétique.

Bien qu'applicables à de multiples situations<sup>89</sup> qui vont au-delà du domaine de la modélisation moléculaire, nous contraindrons notre description au domaine de la modélisation moléculaire dans le cadre de ce chapitre.

## 1. Population initiale

En premier lieu, une population de conformères est générée de manière aléatoire. Chaque conformère est par la suite « encodé » de manière à ce que toute l'information concernant sa géométrie puisse être décrite par une chaîne binaire (une séquence de 0 et de 1 représentant le code génétique de chaque individu). Cet encodage est propre à tous les logiciels, mais réfère habituellement en chimie computationnelle aux valeurs des longueurs de liaisons et angles des molécules mis dans une base spécifique. Par exemple, une distance interatomique de 153 pm peut être représentée en sa valeur binaire par 10 011 001 (équation 4.1) :

$$153 \text{ pm} = 1 \times 2^7 + 0 \times 2^6 + 0 \times 2^5 + 1 \times 2^4 + 1 \times 2^3 + 0 \times 2^2 + 0 \times 2^1 + 1 \times 2^0 \quad (4.1)$$

## 2. Évaluation

Cette étape consiste à qualifier chaque conformation en lui attribuant un score. Ce score est déterminé en calculant, dans le cas qui nous intéresse, l'énergie d'interaction entre le ligand et le récepteur. Plus l'énergie d'interaction est favorable à l'arrimage, plus le score est élevé. Une description plus élaborée de la fonction de score que nous utiliserons sera présentée dans la méthodologie.

## 3. Sélection

Un certain nombre de conformères est choisi de manière aléatoire. Plus le score du conformère est élevé, plus la probabilité que le conformère soit choisi est grande.

## 4. Croisements et mutations

De manière aléatoire, les conformères sélectionnés sont croisés entre eux. C'est à cette étape que se révèle la nécessité de transformer l'information géométrique des conformères en code binaire. En effet, pour croiser les conformères sélectionnés, il ne suffit que de

prendre une partie de la séquence binaire du conformère **A** de manière aléatoire avec la partie complémentaire du conformère **B**. En combinant les deux parties, on obtient une nouvelle séquence qui correspond à un nouveau conformère **C**. On peut *a priori* penser que ce croisement permettra de coupler les caractéristiques qui ont donné les scores élevés des conformères **A** et **B**. En pratique, si on croise les éléments favorisant une énergie favorisant l'arrimage, le nouveau conformère **C** aura un score élevé lors de la prochaine phase d'évaluation. En revanche, si les éléments défavorisant l'arrimage de **A** et **B** sont retenus dans **C**, le score de **C** sera faible lors de la prochaine phase d'évaluation et il y aura de fortes chances que ce conformère ne soit pas sélectionné par la suite. On peut ainsi voir qu'à mesure que les itérations ont lieu, on favorisera toujours les meilleures conformères. Cette étape est réalisée à de multiples reprises pour donner une nouvelle population de conformères. Conjointement aux croisements, on peut procéder à des mutations des conformères. De manière aléatoire, on permutera certains 0 et 1 pour obtenir de nouveaux conformères qui n'auraient pu exister autrement.

## 5. Critère d'arrêt

Cette dernière étape consiste à décider si une autre itération est nécessaire. La plupart du temps, le nombre maximal d'itérations est décidé de manière plus ou moins arbitraire. Dans certains cas, on peut aussi décider d'arrêter les itérations lorsque les conformères se ressemblent trop. Si le critère d'arrêt n'est pas atteint, l'étape 2 est reprise avec la nouvelle population obtenue.

L'arrimage moléculaire utilise cette méthodologie pour évaluer les meilleurs agencements entre une protéine et un ligand. Il considère la protéine fixe et encode les informations structurelles du ligand (position, longueur de liaison, angle, etc.) Il génère une population de ligands de manière aléatoire puis utilise une fonction de score pour les qualifier les uns par rapport aux autres. Par la suite, une sélection est faite et des croisements et des

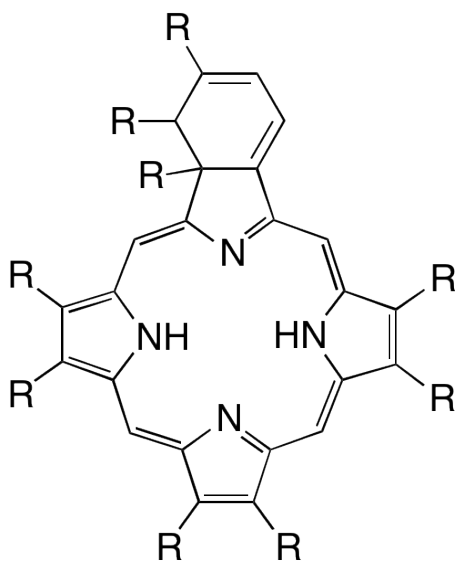
mutations sont réalisés pour conserver certaines caractéristiques structurales des conformères favorisant une basse énergie d'interaction ligand-récepteur. Ce processus est répété de manière itérative de sorte à toujours favoriser les meilleures conformations pour ultimement arriver au meilleur arrimage du ligand sur la protéine, sans passer par une évaluation systématique de la surface d'énergie potentielle. Plusieurs logiciels permettent de réaliser des arrimages moléculaires sur des protéines. Si l'algorithme génétique est de manière générale, très semblable, la principale différence entre ces logiciels est la manière dont le score est attribué et les considérations de la fonction de score. Ces particularités seront présentées plus en détail dans la sous-section 4.3.2.

Cette méthodologie de résolution s'applique bien au sujet que nous allons aborder dans ce chapitre : soit l'observation des interactions supramoléculaires sur de très grands systèmes. Plus particulièrement, nous restreindrons ce sujet à l'observation des interactions entre une protéine particulière et une série de candidats organiques pouvant jouer le rôle de ligands.

#### **4.2. Contexte de l'étude**

Cette étude a été amorcée par une société spécialisée en recherche clinique qui avait besoin d'information computationnelle sur leurs systèmes biologiques. Étant donné l'aspect confidentiel des résultats découlant de l'ensemble de l'étude, le nom de la société, le nom des inhibiteurs potentiels et le nom, la fonction ainsi que certaines parties de la protéine étudiée ne seront pas divulgués.

La recherche du groupe avait montré qu'un dérivé benzoporphyrine (Figure 4.3) montrait une grande affinité avec une protéine exprimée dans certaines conditions biologiques. Cette affinité se manifestait par une forte inhibition d'un ligand connu de la protéine.



**4.1**

Figure 4.3 – Dérivé de benzoporphyrine **4.1**.

Pour comprendre les raisons de cette affinité, il était question de trouver le(s) site(s) d'arrimage du composé sur la protéine, d'effectuer un criblage virtuel de milliers de composés similaires et d'analyser les bons et les mauvais candidats pour suggérer les facteurs favorisant l'arrimage.

### **4.3. Méthodologie**

La stratégie mise en place au cours de cette étude consistait à mettre en relation certains résultats d'arrimage avec les données expérimentales disponibles. Nous espérions ainsi voir une corrélation entre les résultats computationnels et expérimentaux, nous permettant de voir si l'arrimage se faisait naturellement dans la bonne zone d'arrimage. La zone d'arrimage trouvée, la dernière étape consistait à faire un criblage virtuel de milliers de composés et d'observer les bons et les moins bons candidats pour tenter de trouver les éléments à considérer et à éliminer dans la structure de l'inhibiteur.

#### **4.3.1. Structure de la protéine**

Quelques structures cristallines de la protéine d'intérêt existent et sont disponibles sur la RCSB Protein Data Bank<sup>90</sup>. Une structure cristallisée avec un ligand a été retenue aux fins d'analyse. Le ligand et les molécules de solvant ont été retirés et la protéine laissée telle qu'elle pour l'arrimage. Étant donné que l'information disponible sur cette protéine était peu étoffée, la structure moléculaire de la protéine a été maintenue fixe et seuls les mouvements des ligands ont été considérés.

#### **4.3.2. Fonction de score et FITTED**

Une multitude de logiciels sont disponibles pour trouver les conformères par arrimage moléculaire. Le choix du logiciel passe notamment par l'évaluation de sa fonction de score. Les fonctions de score actuellement développées se classent parmi trois catégories distinctes<sup>91</sup> : les fonctions de score basées sur les champs de force, les fonctions de score empiriques et les fonctions de score de type potentiels-statistiques.

Comme discuté dans les précédents chapitres (équation 1.16), les fonctions de score basées sur les champs de force sont des modèles mathématiques faisant intervenir les interactions interatomiques. Bien qu'ayant un sens physique propre, le défi majeur de ce type de fonction consiste à inclure la contribution du solvant dans le processus d'arrimage. La solution triviale est d'inclure les molécules de solvant de manière explicite, mais cette action est trop coûteuse en temps de calcul pour permettre la mise en œuvre dans des processus de criblage virtuel. Les logiciels DOCK<sup>92</sup>, Autodock<sup>93</sup> et GOLD<sup>94</sup> utilisent cette approche pour qualifier les conformères.

Les fonctions de score empiriques estiment quant à elles l'affinité d'un ligand pour un récepteur en utilisant des termes normalisés caractérisant certains phénomènes tels que l'énergie d'interaction de van der Waals, les interactions électrostatiques, l'entropie, l'hydrophobie et d'autres. La validation de la fonction de score est réalisée à partir de complexes ligand-protéines connus. Étant donné la simplification des termes d'énergie, les algorithmes utilisant ce type de fonction de score sont beaucoup plus rapides que celles utilisant les fonctions de score basées sur les champs de force<sup>91</sup>, ce qui en fait une méthode de choix pour les criblages virtuels. FlexX<sup>95</sup>, Glide<sup>96</sup> et FITTED<sup>97</sup> font partie des logiciels utilisant cette classe de fonctions de score.

Le dernier type de fonction de score utilise des potentiels énergétiques dérivés de l'information expérimentale disponible. Les potentiels sont obtenus en calculant la fréquence d'occurrence des paires d'atomes en utilisant la relation inverse de Boltzmann<sup>98</sup> (équation 4.2).

$$w(r) = -k_B T \times \ln \left[ \frac{\rho(r)}{\rho^*(r)} \right] \quad (4.2)$$

où  $w(r)$  est le potentiel de la paire,  $k_B$  est la constante de Boltzmann,  $\rho(r)$  est la probabilité d'observation des paires d'atomes protéines-ligand à une distance  $r$  et  $\rho^*(r)$  est la probabilité d'un état de référence où les interactions interatomiques sont nulles. En comparaison par rapport aux deux autres catégories de fonctions, les fonctions de type potentiels-statistiques offrent une bonne balance entre la précision et la vitesse de calcul. En effet, étant donné que les paramètres découlent de bases de données larges et diverses, cette méthode a l'avantage d'être particulièrement robuste et insensible au type de données analysées<sup>99</sup>. L'obstacle principal consiste néanmoins à déterminer l'état de référence



$\rho^*(r)$ <sup>91</sup>. MScore<sup>100</sup> et GOLD/ASP<sup>101</sup> font partie des logiciels utilisant cette dernière classe de fonction de score.

FITTED a été utilisé pour toutes les étapes, de l'arrimage local au criblage virtuel. Ce choix a notamment été motivé par le comparatif<sup>102</sup> fait avec les autres logiciels d'arrimage principaux qui avantageait FITTED dans les expériences d'arrimage croisé, où le ligand arrimé ne fait pas initialement partie de la structure cristallisée (Figure 4.4).

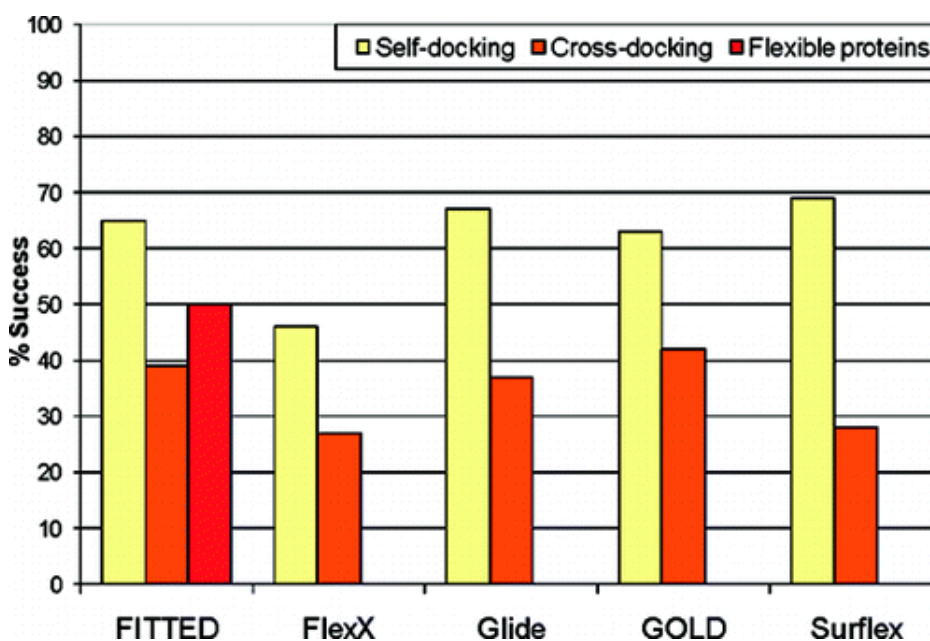


Figure 4.4 – Pourcentage de succès de différentes expériences d'arrimage avec plusieurs logiciels<sup>102</sup>.

On peut voir que de tous les logiciels d'arrimage testés, FITTED donne des résultats supérieurs ou comparables aux logiciels analogues, en particulier dans le contexte de l'arrimage croisé, où le ligand arrimé n'était pas initialement cristallisé dans la protéine. Pour comprendre ce qui faisait de FITTED un logiciel plus performant, nous nous sommes attardés à l'étude de sa fonction de score.

Dans le cas qui nous intéresse, la fonction de score consistait en une évaluation empirique des changements entropiques, des ponts hydrogène, des interactions électrostatiques et van der Waals et de la contribution de l'énergie libre de liaison<sup>103</sup>. La fonction de score de FITTED prend la forme suivante :

$$\begin{aligned}\Delta G_{\text{liaison}} = & \Delta G_0 + 0,14N_{ROT} \\ & + \sum (\text{facteur}) [0,26U_{vdW}^{inh-prot} + 0,035U_{elec}^{inh-prot} \\ & + 0,80f_{hb}(\Delta r, \Delta \alpha)]\end{aligned}\quad (4.3)$$

où  $\Delta G_{\text{liaison}}$  représente l'énergie libre de liaison protéine-ligand et  $\Delta G_0$  représente une contribution à l'énergie de liaison qui ne dépend pas d'une interaction spécifique avec le récepteur. Cette dernière peut être considérée comme une réduction de l'énergie de liaison causée par la perte d'entropie de translation et de rotation du ligand<sup>104</sup>. Ce terme permet d'ailleurs d'obtenir un gain net de la fonction de score<sup>103</sup> à discréditer les composés actifs des composés inactifs.  $N_{ROT}$  représente le nombre de liaisons acycliques  $sp^3$ - $sp^3$  et  $sp^3$ - $sp^2$ . Ce terme permet de tenir compte de la différence d'énergie de liaison engendrée par la perte de degrés de liberté internes du ligand. Les termes  $U_{vdW}^{inh-prot}$  et  $U_{elec}^{inh-prot}$  se basent sur le calcul d'énergie libre impliquant les interactions van der Waals et électrostatiques entre le ligand et le récepteur d'après les informations du champ de force implémenté dans le logiciel. Ce calcul est réalisé en observant la différence énergétique entre le ligand lié et non lié au récepteur<sup>105</sup>. Finalement,  $f_{hb}(\Delta r, \Delta \alpha)$  est une fonction permettant d'évaluer les ponts hydrogène en fonction de la longueur ( $\Delta r$ ) et de l'angle du pont H ( $\Delta \alpha$ ) (ex.  $O \cdots H-N = 180^\circ$ )<sup>103</sup>.

### **4.3.3. Choix des molécules cibles pour le criblage virtuel**

Les molécules cibles du criblage virtuel ont été obtenues en consultant la base de données ZINC<sup>106</sup>. À partir de toutes les molécules disponibles, les règles de Lipinski<sup>107</sup> ont été appliquées pour éliminer les composés qui ne feraient pas de bons candidats pour la préparation d'un médicament pris par voie orale. Les règles sont décrites ci-dessous :

1. Ne possède pas plus de 5 donneurs de ponts hydrogène.
2. Ne possède pas plus de 10 accepteurs de ponts hydrogène.
3. Possède une masse moléculaire inférieure à 500 daltons.
4. Possède un log P inférieur ou égal à 5 pour le coefficient de partition octanol-eau.

Une sélection aléatoire des composés restants a été faite pour restreindre le criblage à environ 4000 composés.

### **4.1. Sommaire des résultats expérimentaux**

Les premières analyses expérimentales se résument à réaliser des études comparatives sur différents candidats analogues à la benzoporphyrine **4.1**. L'expérience consistait à observer les constantes de liaison d'un anticorps à la protéine en présence de chaque candidat.

L'hypothèse initiale étant que le site de liaison des candidats était situé assez près de l'épitope de la protéine, l'inhibition de la constante de liaison de l'anticorps à la protéine était un indice que l'environnement local de l'épitope avait été suffisamment modifié pour que l'anticorps ne reconnaisse plus la protéine. Cette observation permettait donc de qualifier les candidats en fonction de la capacité de l'anticorps à se fixer à la protéine. Plus les constantes de liaisons entre l'anticorps et la protéine étaient faibles, plus l'interaction

entre le candidat et la protéine était forte. Nous avons utilisé par conséquent ces résultats préliminaires pour tenter de valider notre modèle mathématique.

#### 4.2. Validation du modèle

Nous avons réalisé l'arrimage moléculaire de la benzoporphyrine **4.1** et de sept autres candidats similaires préalablement testés *in vitro*. Par la suite, nous avons comparé les scores obtenus avec les constantes de liaison de l'anticorps à l'antigène (Figure 4.5).

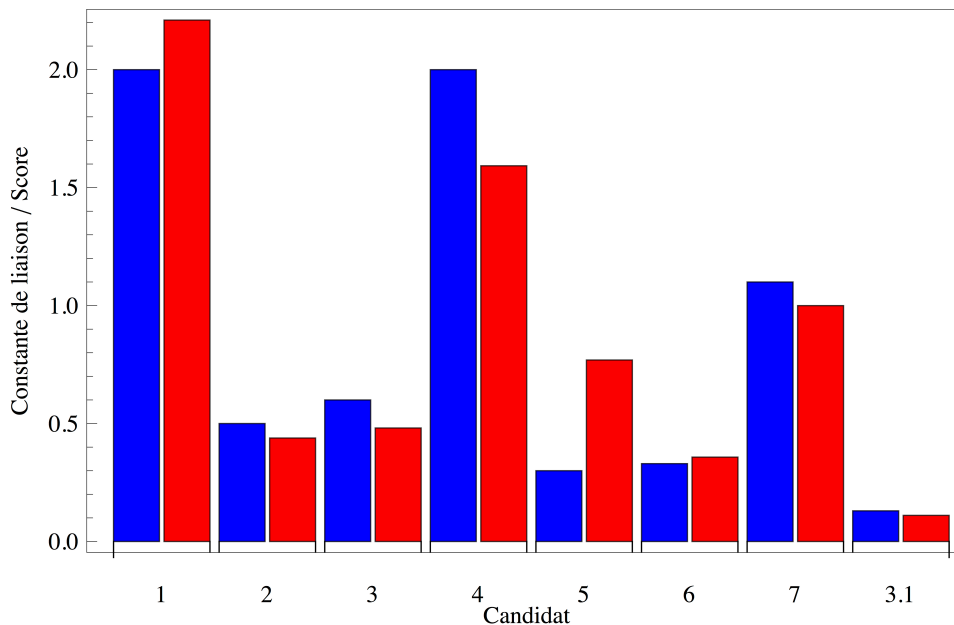


Figure 4.5 – Comparaison des constantes de liaison de l'anticorps avec la protéine en présence des différents candidats (en bleu) avec les scores obtenus par l'arrimage (en rouge).

Afin de faire correspondre visuellement les scores aux données expérimentales, nous avons déduit la relation mathématique suivante :

$$\gamma = -\frac{155,6}{\sigma} - 2,112 \quad (4.4)$$

où  $\sigma$  représente le score obtenu et  $\gamma$  la constante de liaison correspondante. Une corrélation de 0,94 a été obtenue.

La comparaison des résultats computationnels avec les constantes de liaison confirme que la fonction de score permet de qualifier assez fidèlement les candidats entre eux. En effet, on peut voir que les candidats qui interagissent le moins fortement avec la protéine (candidats 1, 4 et 7) possèdent également les moins bons scores. Conséquemment, les meilleurs candidats observés expérimentalement ont en général un meilleur score. Finalement, tel qu'illustré par les observations expérimentales, le meilleur candidat demeure la benzoporphyrine **4.1**. Ces résultats nous ont encouragés à poursuivre nos calculs afin de déterminer la zone d'arrimage.

### 4.3. Arrimage moléculaire

Un arrimage global a été réalisé afin de déterminer la zone où se fixent les candidats sur la protéine. La Figure 4.6 montre la zone d'arrimage du composé **4.1**.

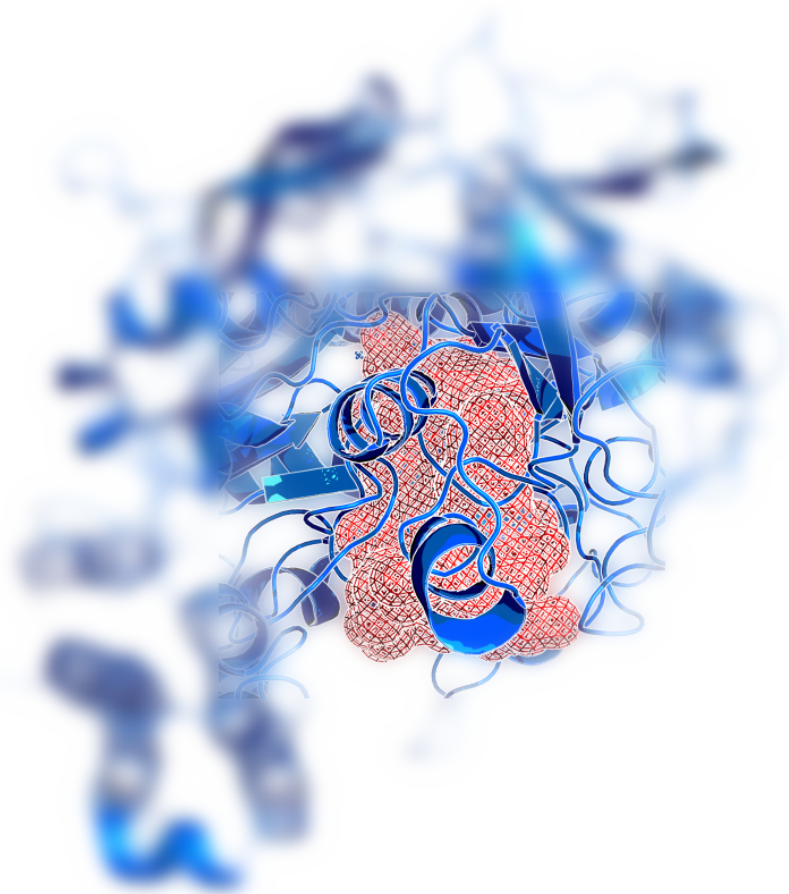


Figure 4.6 – Zone d'arrimage de la benzoporphyrine **4.1** dans la protéine.

On constate que la zone d'arrimage couvre une cavité qui se rend jusqu'au cœur de la protéine. Les candidats étant des dérivés porphyrines volumineux, il est donc tout à fait plausible que la déformation résultante de l'interaction entre nos candidats et la protéine empêche l'anticorps de reconnaître l'épitope, phénomène observé lors des essais de constante de liaison entre l'anticorps et la protéine. Ces résultats nous ont permis d'amorcer l'étape du criblage virtuel.

#### 4.4. Criblage virtuel

Le criblage virtuel consistait à réaliser l'arrimage moléculaire d'une multitude de composés dans la zone d'arrimage précédemment illustrée afin de trouver de nouveaux candidats potentiels. Cette étape permettait aussi d'avoir un regard nouveau sur les sites d'interaction les plus importants dans le processus d'interaction ligand-récepteur afin de dresser des caractéristiques importantes de l'inhibiteur. Tel que mentionné dans la section 4.3.3, environ 4000 molécules choisies parmi des candidats plausibles ont été analysées.

Les meilleurs candidats ont été sélectionnés en fonction de leur score. La Figure 4.7 montre les meilleurs candidats en comparant les valeurs  $\gamma$  calculées selon l'équation 3.4. Par souci de clarté, l'inverse des valeurs  $\gamma$  est illustré. On peut remarquer que la benzoporphyrine **4.1** demeure dans les meilleurs candidats. Afin de voir les éléments nuisant aux interactions entre les candidats et la protéine, les plus grandes valeurs  $\gamma$  ont été sélectionnées. La Figure 4.8 montre les candidats sélectionnés.

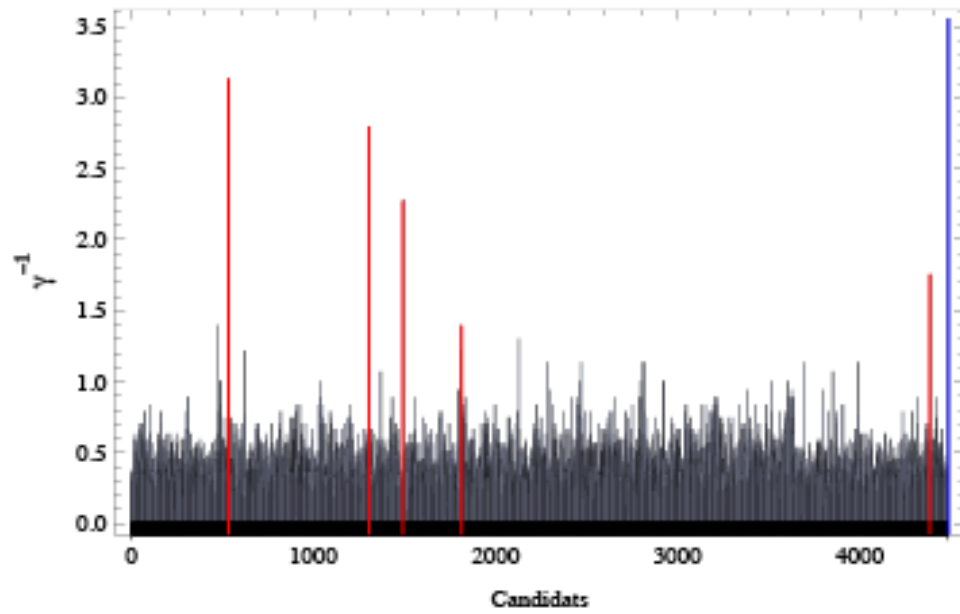


Figure 4.7 – Inverse des valeurs  $\gamma$  pour tous les candidats. Les barres rouges réfèrent aux meilleurs candidats. La barre bleue réfère au candidat **4.1**.



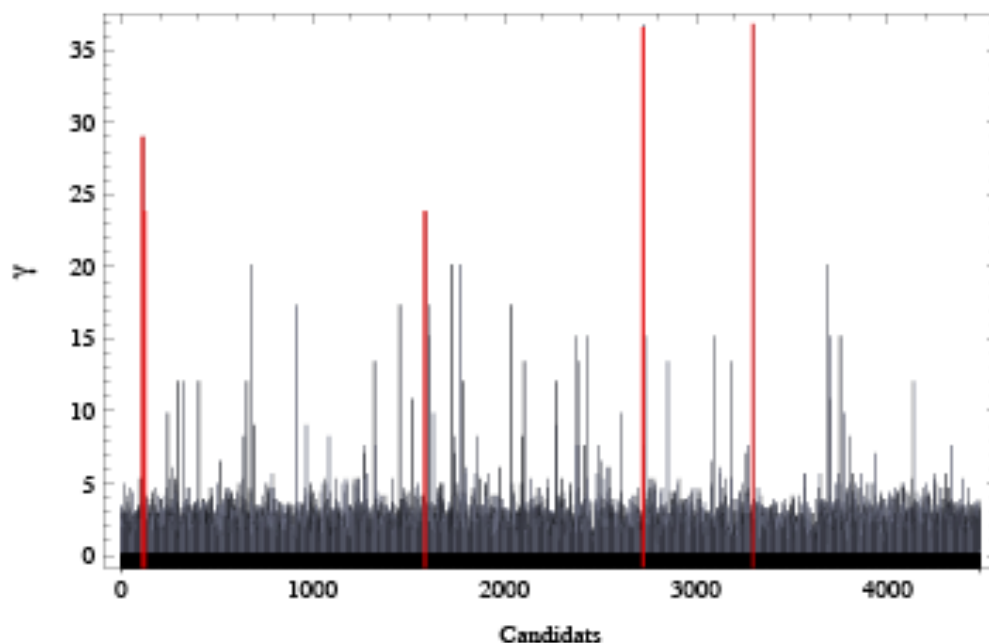


Figure 4.8 – Valeurs  $\gamma$  pour tous les candidats. Les barres rouges réfèrent aux mauvais candidats. Le candidat **4.1** n'est pas illustré puisqu'il fait partie du bruit entre 0 et 5.

L'observation des deux catégories de candidats a permis d'extraire trois caractéristiques structurales essentielles d'un bon candidat : il devait posséder un groupement chargé, il devait posséder un cœur aromatique et ces deux groupements devaient être séparés par une certaine distance.

### 1. Groupement chargé

La Figure 4.9 montre une superposition des candidats possédant les meilleures valeurs  $\gamma$ . Étant donné la présence d'un cation au fond de la cavité d'arrimage, la présence d'un groupement chargé stabilisant le cation est essentielle à un arrimage efficace. Tous nos meilleurs candidats possédaient d'ailleurs ce type de groupement, qu'il s'agisse de groupements nitro ou carboxylates. Cette caractéristique ne garantissait toutefois pas un

bon candidat. En effet, des groupements chargés ont été observés même sur certains des mauvais candidats (Figure 4.10).

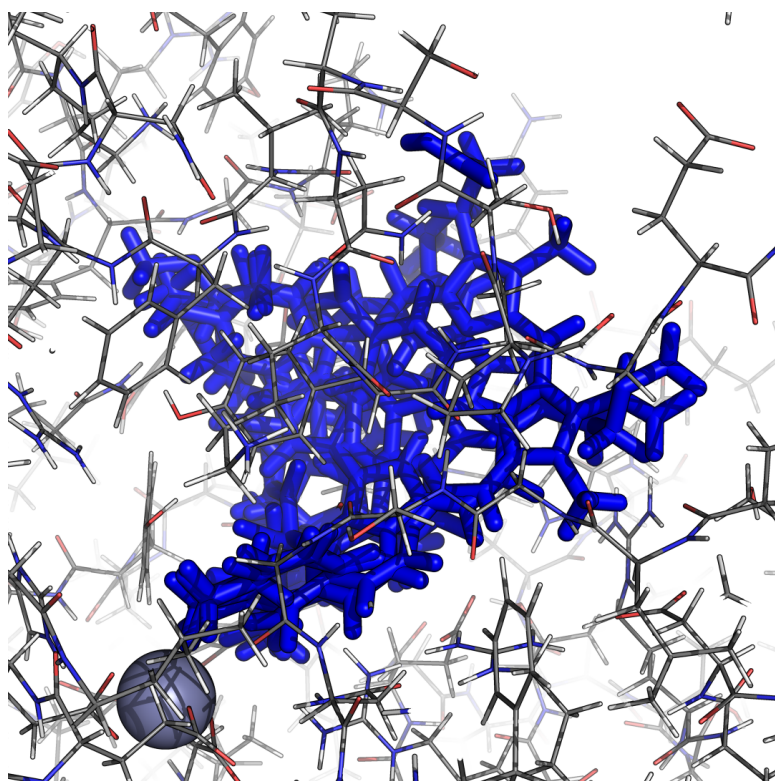


Figure 4.9 – Arrimage des meilleurs candidats. Les candidats sont en bleu alors que la protéine est illustrée selon les couleurs usuelles des représentations atomiques. Le cation est illustré par une sphère.

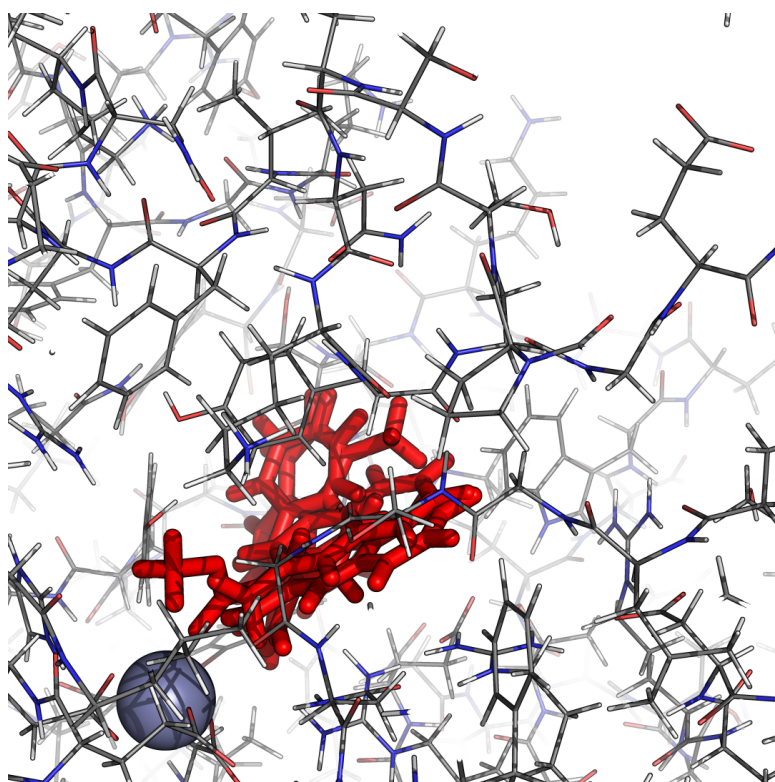


Figure 4.10 – Arrimage des mauvais candidats. Les candidats sont en rouge alors que la protéine est illustrée selon les couleurs usuelles des représentations atomiques. Le cation est illustré par une sphère.

## 2. Cœur aromatique

Tous les bons candidats possèdent une autre caractéristique : celle d'avoir au moins un noyau aromatique. Nous avons pu conclure que cette composante moléculaire permet une intercalation favorisant des interactions  $\pi$ - $\pi$  et CH- $\pi$  entre une tyrosine et un tryptophane de la protéine (Figure 4.11). Par ailleurs, cette intercalation n'est pas aussi clairement observable dans le cas des candidats avec de mauvaises valeurs  $\gamma$ .

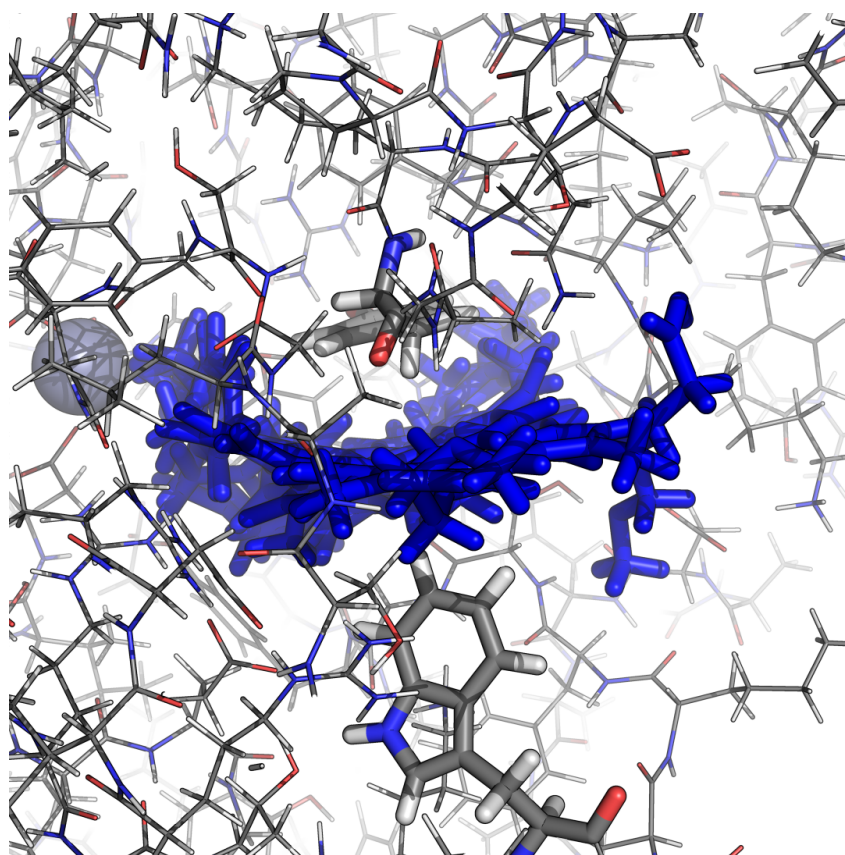


Figure 4.11 – Intercalation des meilleurs candidats. Les candidats sont en bleu alors que la protéine est illustrée selon les couleurs usuelles des représentations atomiques. La tyrosine et le tryptophane sont illustrés en représentation large de part et d'autre des ligands.

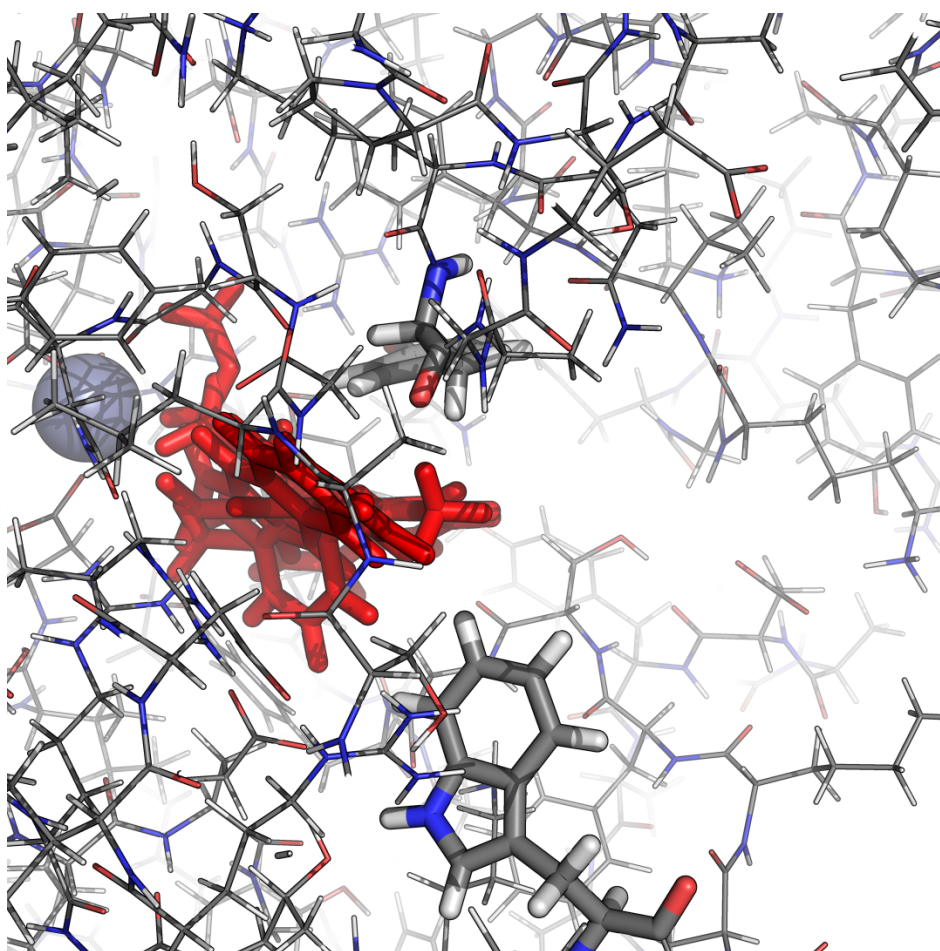


Figure 4.12 – Intercalation des mauvais candidats. Les candidats sont en rouge alors que la protéine est illustrée selon les couleurs usuelles des représentations atomiques. La tyrosine et le tryptophane sont illustrés en représentation large.

Étant donné que même les candidats avec une valeur  $\gamma$  élevée possèdent tout de même des noyaux aromatiques, cette observation nous a conduit à conclure que la distance entre ces caractéristiques moléculaires est primordiale.

### 3. Distance entre les groupements

Le calcul de la distance approximative entre les groupements fonctionnels montre l'importance de la position des fonctions sur la molécule. En effet, le Tableau VII illustre

les différentes distances trouvées et montre comment elles permettent une discrimination entre les bons et les mauvais candidats.

Tableau VII – Comparaison entre les valeurs  $\gamma$  et les distances entre les groupements fonctionnels.

	Valeur $\gamma$	Distance (Å)	Distance moyenne (Å)
Mauvais candidats	29	4,1	$4,3 \pm 0,3$
	23	4,1	
	23	4,5	
	20	4,2	
	20	4,8	
Bons candidats	0,72	9,0	$8,4 \pm 0,9$
	0,57	8,7	
	0,44	9,0	
	0,36	7,1	
	0,31	8,0	
<b>4.1</b>	0,13	9,7	9,7

#### **4.5. Limites et conclusions**

En somme, l'arrimage moléculaire a été utilisé afin de reproduire les résultats expérimentaux et tenter d'orienter les recherches subséquentes à la conception d'un nouvel inhibiteur enzymatique. Nous avons en premier lieu validé l'usage de la fonction de score en comparant les résultats expérimentaux disponibles avec les résultats de la fonction de score des arrimages. Cette validation a été confirmée par une corrélation appréciable entre les deux séries de données.

Ces résultats nous ont encouragés à poursuivre nos analyses et à trouver la zone d'arrimage des candidats. La zone trouvée, nous avons été en mesure d'effectuer un criblage virtuel de plus de 4000 candidats potentiels. Les résultats obtenus montrent que les bons candidats doivent posséder trois caractéristiques structurales principales : ils doivent posséder un groupement nucléophile pour stabiliser la charge sur la protéine, ils doivent posséder un centre aromatique pour favoriser l'intercalation et assurer la pénétration dans la zone d'arrimage et finalement, ces groupements doivent être suffisamment espacés pour que leur fonction respective puisse être accomplie.

La méthodologie utilisée est toutefois contrainte par certaines limites qui nécessiteraient une validation expérimentale subséquente. En effet, l'arrimage a été réalisé sur une protéine complètement rigide. Bien que FITTED gère d'une certaine façon les mouvements des récepteurs, l'information dont nous disposions sur la protéine n'était pas assez étoffée pour être en mesure de mettre en place une méthodologie impliquant une protéine ayant des acides aminés mobiles. Par conséquent, l'arrimage a été effectué avec une protéine fixée dans une conformation où un ligand interagissait déjà avec la protéine.

Un désavantage d'utiliser des méthodes approximatives pour simuler des observables est que les conclusions découlant de ces méthodes nécessitent l'implication de résultats

expérimentaux pour les supporter. Dans le cas qui nous intéresse, des preuves expérimentales que les candidats inhibent plus ou moins bien seraient nécessaires à la confirmation des conclusions apportées dans ce chapitre.



## Chapitre 5: Conclusion

Tout au long de cette étude, nous nous sommes questionné sur l'applicabilité de différents modèles mathématiques en chimie. Nous avons testé cette approche sur plusieurs systèmes en nous assurant d'une rapidité de calcul qui n'allait pas au détriment de la précision de la description de l'objet modélisé.

En débutant avec des systèmes moléculaires très simples et des modèles quantiques assidus, nous avons, au cours des chapitres subséquents, étudié des objets chimiques de plus en plus complexes en nous assurant de formuler des approximations qui, bien qu'imposant leurs limites, permettaient de répondre à nos questions tout en diminuant le poids computationnel qu'auraient imposé les méthodologies plus rigoureuses.

De manière générale, la DFT dépendante du temps avec la fonctionnelle PBE0 s'est avérée très efficace pour étudier des composés photoactifs et prédire correctement les énergies d'excitation de ces molécules. Par ailleurs, en ne considérant que les états stationnaires et les orbitales naturelles d'un système moléculaire analogue au 1,3,5-triazapentadiène, nous avons pu mieux comprendre et rationaliser différents effets stabilisant l'état de transition des mésomères de ce type de composé.

Le remplacement de la fonction d'onde multiélectronique par une fonction empirique basée sur l'additivité de potentiels atomiques a permis de simuler de plus gros systèmes tout en émulant le caractère dynamique des ensembles moléculaires. L'utilisation de la dynamique moléculaire et du champ de force GAFF a permis de simuler l'encapsulation de pyrène par un trimère d'un dérivé d'acide cholique et les interactions d'un rotaxane-parapluie à l'interface d'une séparation de phase entre l'eau et le dichlorométhane. Une meilleure compréhension des mécanismes d'agrégation dans divers solvants est ressortie de ces modélisations.

Finalement, l'utilisation d'algorithmes génétiques et de la fonction de score empirique implémentée dans FITTED a permis d'évaluer, à faible coût, le potentiel d'interaction entre une protéine particulière et des milliers de molécules candidates. Cette comparaison systématique a permis de suggérer certaines prédispositions structurales à considérer dans la conception d'un nouvel inhibiteur enzymatique.

Globalement, la modélisation moléculaire a pu répondre de manière partielle ou complète aux questions posées sur chaque système moléculaire. Il a été montré qu'un compromis peut être fait entre la rapidité des calculs, la taille des systèmes et la validité des résultats. Par ailleurs, cette étude met l'accent sur le fait qu'il n'existe pas un seul bon modèle et une seule méthodologie applicable à tout système, et qu'une connaissance adéquate des équilibres moléculaires propre à chaque système est critique à l'élaboration d'une simulation judicieuse. Ce mémoire adresse finalement un regard global sur certaines questions fondamentales à considérer lorsqu'une méthodologie de modélisation moléculaire est développée.

Avec les constantes avancées technologiques réalisées tant au niveau des architectures informatiques que de l'optimisation des algorithmes logiciels, il devient incontestable que la place de la modélisation moléculaire ne cessera de croître au cours des années à venir. Il va sans dire que la démocratisation des outils et la création de sociétés de développement sérieuses contribueront à l'essor de cette facette de la chimie. Nous croyons que cette étude s'inscrit dans une optique de guide sur les choix à faire face à certains systèmes de tailles et de complexité variables.

## Bibliographie

1. Gelilei, G., *The Assayer*. University of Pennsylvania Press: Pennsylvania, 1960.
2. Eykhoff, P., *System identification: Parameter and state estimation*. Wiley-Interscience: London, 1974; p 555.
3. (a) Beer, F. P., *Mécanique pour ingénieurs*. McGraw-Hill: Montréal, 1998; Vol. 1;  
(b) Jones, J. C.; Ertas, A., *The engineering design process*. John Wiley & Sons: New York, 1993.
4. McQuarrie, D. A., *Quantum Chemistry*. University Science Books: Sausalito (CA), 1983.
5. Rutherford, E., The Structure of the Atom. *Philosophical Magazine* **1911**, 21 (6), 669-688.
6. Schrödinger, E., An Undulatory Theory of the Mechanics of Atoms and Molecules. *Physical Review* **1926**, 28 (6), 1049-1070.
7. (a) Born, M.; Oppenheimer, R., Zur Quantentheorie der Molekeln. *Annalen der Physik* **1927**, 389 (20), 457-484; (b) Blinder, S. M. On the Quantum Theory of Molecules. <http://www.ulb.ac.be/cpm/people/scientists/bsutclif/bornop.pdf>.
8. Jensen, F., *Introduction to Computational Chemistry*. John Wiley & Sons: Chichester, England, 1999.
9. Dewar, M. J. S.; Thiel, W., Ground states of molecules. 38. The MNDO method. Approximations and parameters. *Journal of the American Chemical Society* **1977**, 99 (15), 4899-4907.

10. Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P., Development and use of quantum mechanical molecular models. 76. AM1: a new general purpose quantum mechanical molecular model. *Journal of the American Chemical Society* **1993**, *115* (12), 5348-5348.
11. Pople, J. A.; Segal, G. A., Approximate Self-Consistent Molecular Orbital Theory. II. Calculations with Complete Neglect of Differential Overlap. *The Journal of Chemical Physics* **1965**, *43* (10), S136-S151.
12. Beveridge, D. L.; Dobosh, P. A., Spin Contamination in Spin-Unrestricted INDO Molecular-Orbital Wavefunctions. *Journal of Chemical Physics* **1968**, *48* (12), 5532-5536.
13. Koch, W.; Holthausen, M. C., *A Chemist's Guide to Density Functional Theory*. second edition ed.; Wiley-Vch: Weinheim, 2001.
14. Kohn, W.; Sham, L. J., Self-Consistent Equations Including Exchange and Correlation Effects. *Physical Review* **1965**, *140* (4A), A1133-A1138.
15. Hohenberg, P.; Kohn, W., Inhomogeneous Electron Gas. *Physical Review* **1964**, *136* (3B), B864-B871.
16. Ceperley, D. M.; Alder, B. J., Ground State of the Electron Gas by a Stochastic Method. *Physical Review Letters* **1980**, *45* (7), 566-569.
17. Truchon, J.-F. Modéliser la polarisation électronique par un continuum diélectrique intramoléculaire. Vers un champ de force polarisable pour la chimie bioorganique. Université de Montréal, Montréal, 2008.
18. Jones, J. E., On the determination of Molecular Fields. II. From the Equation of State of a Gas. *Proceedings of the Royal Society A* **1924**, *106* (738), 463-477.
19. Onofrio, N. Modélisation de l'interaction d'échange par théorie de la fonctionnelle de la densité couplée au formalisme de la symétrie brisée. Université de Grenoble, Grenoble, 2011.

20. (a) Raghavachari, K.; Pople, J. A., Calculation of one-electron properties using limited configuration interaction techniques. *International Journal of Quantum Chemistry* **1981**, 20 (5), 1067-1071; (b) Pople, J. A.; Seeger, R.; Krishnan, R., Variational configuration interaction methods and comparison with perturbation theory. *International Journal of Quantum Chemistry* **1977**, 12 (S11), 149-163; (c) Krishnan, R.; Schlegel, H. B.; Pople, J. A., Derivative studies in configuration--interaction theory. *The Journal of Chemical Physics* **1980**, 72 (8), 4654-4655.
21. (a) Eade, R. H. A.; Robb, M. A., Direct minimization in mc scf theory. the quasi-newton method. *Chemical Physics Letters* **1981**, 83 (2), 362-368; (b) Frisch, M.; Ragazos, I. N.; Robb, M. A.; Schlegel, H. B., An evaluation of three direct MC-SCF procedures. *Chemical Physics Letters* **1992**, 189 (6), 524-528; (c) Yamaguchi, Y.; Frisch, M.; Gaw, J.; Schaefer Iii, H. F.; Binkley, J. S., Analytic evaluation and basis set dependence of intensities of infrared spectra. *The Journal of Chemical Physics* **1986**, 84 (4), 2262-2278.
22. Møller, C.; Plesset, M. S., Note on an Approximation Treatment for Many-Electron Systems. *Physical Review* **1934**, 46 (7), 618-622.
23. (a) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J., Ab Initio Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields. *The Journal of Physical Chemistry* **1994**, 98 (45), 11623-11627; (b) Becke, A. D., Density-functional thermochemistry. III. The role of exact exchange. *The Journal of Chemical Physics* **1993**, 98 (7), 5648-5652.
24. (a) Ernzerhof, M.; Scuseria, G. E., Assessment of the Perdew--Burke--Ernzerhof exchange-correlation functional. *The Journal of Chemical Physics* **1999**, 110 (11), 5029-5036; (b) Adamo, C.; Barone, V., Toward reliable density functional methods without adjustable parameters: The PBE0 model. *The Journal of Chemical Physics* **1999**, 110 (13), 6158-6170.
25. De La Pierre, M.; Orlando, R.; Maschio, L.; Doll, K.; Ugliengo, P.; Dovesi, R., Performance of six functionals (LDA, PBE, PBESOL, B3LYP, PBE0, and WC1LYP) in

the simulation of vibrational and dielectric properties of crystalline compounds. The case of forsterite  $\text{Mg}_2\text{SiO}_4$ . *Journal of Computational Chemistry* **2011**, 32 (91775-1784).

26. Becke, A. D., Density-functional exchange-energy approximation with correct asymptotic behavior. *Physical Review A* **1988**, 38 (6), 3098-3100.

27. Vosko, S. H.; Wilk, L.; Nusair, M., Accurate spin-dependent electron liquid correlation energies for local spin density calculation: a critical analysis. *Canadian Journal of Physics* **1980**, 58 (11), 1200-1211.

28. Lee, C.; Yang, W.; Parr, R. G., Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Physical Review B* **1988**, 37 (2), 785-789.

29. Curtiss, L. A.; Raghavachari, K.; Trucks, G. W.; Pople, J. A., Gaussian-2 theory for molecular energies of first- and second-row compounds. *The Journal of Chemical Physics* **1991**, 94 (11), 7221-7230.

30. Paier, J.; Marsman, M.; Kresse, G., Why does the B3LYP hybrid functional fail for metals? *The Journal of Chemical Physics* **2007**, 127 (2), 024103-10.

31. Perdew, J. P.; Burke, K.; Ernzerhof, M., Generalized Gradient Approximation Made Simple. *Physical Review Letters* **1996**, 77 (18), 3865-3868.

32. (a) Ernzerhof, M., Construction of the adiabatic connection. *Chemical Physics Letters* **1996**, 263, 499-506; (b) Perdew, J. P.; Ernzerhof, M.; Burke, K., Rationale for mixing exact exchange with density functional approximations. *Journal of Chemical Physics* **1996**, 105 (22), 9982-9985; (c) Ernzerhof, M.; Perdew, J. P.; Burke, K., Coupling-constant dependence of atomization energies. *International Journal of Quantum Chemistry* **1997**, 64 (3), 285-295.

33. Harris, J., Adiabatic-connection approach to Kohn-Sham theory. *Physical Review A* **1984**, 29 (4), 1648-1659.

34. (a) Wathelet, V.; Preat, J.; Bouhy, M.; Fontaine, M.; Perpète, E. A.; André, J.-M.; Jacquemin, D., Assessment of PBE0 for evaluating the absorption spectra of carbonyl molecules. *International Journal of Quantum Chemistry* **2006**, *106* (8), 1853-1859; (b) Adamo, C.; Scuseria, G. E.; Barone, V., Accurate excitation energies from time-dependent density functional theory: Assessing the PBE0 model. *The Journal of Chemical Physics* **1999**, *111* (7), 2889-2899; (c) Jacquemin, D.; Perpète, E. A.; Scuseria, G. E.; Ciofini, I.; Adamo, C., TD-DFT Performance for the Visible Absorption Spectra of Organic Dyes: Conventional versus Long-Range Hybrids. *Journal of Chemical Theory and Computation* **2007**, *4* (1), 123-135.
  
35. (a) Petersson, G. A.; Bennett, A.; Tensfeldt, T. G.; Al-Laham, M. A.; Shirley, W. A.; Mantzaris, J., A complete basis set model chemistry. I. The total energies of closed-shell atoms and hydrides of the first-row elements. *The Journal of Chemical Physics* **1988**, *89* (4), 2193-2218; (b) Petersson, G. A.; Al-Laham, M. A., A complete basis set model chemistry. II. Open-shell systems and the total energies of the first-row atoms. *The Journal of Chemical Physics* **1991**, *94* (9), 6081-6090.
  
36. (a) McLean, A. D.; Chandler, G. S., Contracted Gaussian basis sets for molecular calculations. I. Second row atoms,  $Z=11-18$ . *The Journal of Chemical Physics* **1980**, *72* (10), 5639-5648; (b) Krishnan, R.; Binkley, J. S.; Seeger, R.; Pople, J. A., Self-consistent molecular orbital methods. XX. A basis set for correlated wave functions. *The Journal of Chemical Physics* **1980**, *72* (1), 650-654.
  
37. Runge, E.; Gross, E. K. U., Density-Functional Theory for Time-Dependent Systems. *Physical Review Letters* **1984**, *52* (12), 997-1000.
  
38. Gaussian 03, R. C. o. D., M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery, Jr., T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E.

Knox, H. P. Hratchian, J. B. Cross, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez, and J. A. Pople, Gaussian, Inc., Pittsburgh PA, 2003.

39. Hanwell, M.; Curtis, D.; Lonie, D.; Vandermeersch, T.; Zurek, E.; Hutchison, G., Avogadro: An advanced semantic chemical editor, Visualization, and analysis platform. *Journal of Cheminformatics* **2012**, *4* (1), 1758-2946.

40. (a) Peng, C.; Schlegel, H. B., Combining Synchronous Transit and Quasi-Newton Methods to Find Transition States. *Israel Journal of Chemistry* **1993**, *33*, 449-454; (b) Peng, C.; Ayala, P. Y.; Schlegel, H. B.; Frisch, M., Using Redundant Internal Coordinates to Optimize Equilibrium Geometries and Transition States. *Journal of Computational Chemistry* **1996**, *17* (1), 49-56.

41. Reed, A. E.; Weinhold, F., Natural localized molecular orbitals. *The Journal of Chemical Physics* **1985**, *83* (4), 1736-1740.

42. (a) Pottier, N., *Physique statistique hors d'équilibre: Processus irréversibles linéaires*. EDP Sciences: France, 2007; p 524; (b) Chong, D. P., *Recent Advances in Computational Chemistry*. University of British Columbia: Canada, 1995; Vol. 1, p 428.

43. Kaltenbrunner, M.; White, M. S.; Glowacki, E. D.; Sekitani, T.; Someya, T.; Sariciftci, N. S.; Bauer, S., Ultrathin and lightweight organic solar cells with high flexibility. *Nature Communications* **2011**, *3* (770), 1-7.

44. McGehee, M. D.; Topinka, M. A., Solar cells: Pictures from the blended zone. *Nature Materials* **2006**, *5* (9), 675-676.



45. Roncali, J., Molecular engineering of the band gap of pi-conjugated systems. *Macromolecular rapid communications* **2007**, 28 (17), 1761-1775.
46. Fau, C. Y.; Yang, S.-H.; Fau, Y. S.; Hsu, C.-S.; Hsu, C. S., Synthesis of conjugated polymers for organic solar cell applications. *Chem Rev* **2009**, 109 (11), 5868-923.
47. Moore, T. S.; Winmill, T. F., CLXXVII.-The state of amines in aqueous solution. *Journal of the Chemical Society, Transactions* **1912**, 101, 1635-1676.
48. Gilli, G.; Gilli, P., *The nature of the Hydrogen Bond*. Oxford University Press: New York, 2009.
49. (a) Bertolasi, V.; Gilli, P.; Ferretti, V.; Gilli, G., Resonance-Assisted O-H ... O Hydrogen Bonding: Its Role in the Crystalline Self-Recognition of  $\beta$ -Diketone Enols and its Structural and IR Characterization. *Chemistry – A European Journal* **1996**, 2 (8), 925-934; (b) Gilli, G.; Bellucci, F.; Ferretti, V.; Bertolasi, V., Evidence for resonance-assisted hydrogen bonding from crystal-structure correlations on the enol form of the .beta.-diketone fragment. *Journal of the American Chemical Society* **1989**, 111 (3), 1023-1028; (c) Gilli, P.; Gilli, G., Hydrogen bond models and theories: The dual hydrogen bond model and its consequences. *Journal of Molecular Structure* **2010**, 972 (1,Ä3), 2-10.
50. Schlick, T., *Molecular Modeling and Simulation*. 1 ed.; Springer: New York, 2002.
51. Zwanzig, R.; Szabo, A.; Bagchi, B., Levinthal's paradox. *Proceedings of the National Academy of Sciences of the United States of America* **1992**, 89 (1), 20-22.
52. Allinger, N. L.; Sprague, J. T., Conformational analysis. XC. Calculation of the structures of hydrocarbons containing delocalized electronic systems by the molecular mechanics method. *Journal of the American Chemical Society* **1973**, 95 (12), 3893-3907.
53. McQuarrie, D. A., *Statistical Mechanics*. Harper & Row: New York, 1976.
54. Atkins, P. W.; De Paula, J., *Chimie Physique*. 2 ed.; de Boeck: Bruxelles, 2004.

55. Metropolis, N.; Ulam, S., The Monte Carlo Method. *Journal of the American Statistical Association* **1949**, *44* (247), 335-341.
56. Halgren, T. A., Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94. *Journal of Computational Chemistry* **1996**, *17* (5-6), 490-519.
57. Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; Mackerell, A. D., CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *Journal of Computational Chemistry* **2010**, *31* (4), 671-690.
58. Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M., CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *Journal of Computational Chemistry* **1983**, *4* (2), 187-217.
59. Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J., Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *Journal of the American Chemical Society* **1996**, *118* (45), 11225-11236.
60. Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A., Development and testing of a general amber force field. *Journal of Computational Chemistry* **2004**, *25* (9), 1157-1174.
61. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A., A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *Journal of the American Chemical Society* **1995**, *117* (19), 5179-5197.
62. Clark, M.; Cramer III, R. D.; Van Opdenbosch, N., Validation of the general purpose tripos 5.2 force field. *Journal of Computational Chemistry* **1989**, *10* (8), 982-1012.

63. Mayo, S. L.; Olafson, B. D.; Goddard III, W. A., DREIDING: A Generic Force Field for Molecular Simulations. *Journal of Physical Chemistry* **1990**, *94* (26), 8897-8909.
64. Kaminski, G.; Jorgensen, W. L., Performance of the AMBER94, MMFF94, and OPLS-AA Force Fields for Modeling Organic Liquids. *The Journal of Physical Chemistry* **1996**, *100* (46), 18010-18013.
65. Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Antechamber, An Accessory Software Package For Molecular Mechanical Calculations. <http://ambermd.org/antechamber/antechamber.pdf>.
66. Lindahl, E.; Hess, B.; van der Spoel, D., GROMACS 3.0: a package for molecular simulation and trajectory analysis. *Journal of Molecular Modeling* **2001**, *7* (8), 306-317.
67. <http://folding.stanford.edu/>.
68. Stewart, J. J. P. MOPAC2009. <http://openmopac.net>.
69. Rocha, G. B.; Freire, R. O.; Simas, A. M.; Stewart, J. J. P., RM1: A reparameterization of AM1 for H, C, N, O, P, S, F, Cl, Br, and I. *Journal of Computational Chemistry* **2006**, *27* (10), 1101-1111.
70. Stewart, J. J. P., Optimization of parameters for semiempirical methods I. Method. *Journal of Computational Chemistry* **1989**, *10* (2), 209-220.
71. Darden, T.; York, D.; Pedersen, L., Particle mesh Ewald: An  $N \cdot \log(N)$  method for Ewald sums in large systems. *The Journal of Chemical Physics* **1993**, *98* (12), 10089-10092.
72. Debye, P., Näherungsformeln für die Zylinderfunktionen für große Werte des Arguments und unbeschränkt veränderliche Werte des Index. *Mathematische Annalen* **1909**, *67* (4), 535-558.

73. Hestenes R., M.; Stiefel, E., Methods of Conjugate Gradients for Solving Linear Systems. *Journal of Research of the National Bureau of Standards* **1952**, 49 (6), 409 - 436.
74. Hockney, R. W.; Eastwood, J. W., *Computer Simulation Using Particles*. McGraw-Hill: New York, 1981.
75. Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R., Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics* **1984**, 81 (8), 3684-3690.
76. Hess, B.; van der Spoel, D.; Lindahl, E. GROMACS User Manual Version 4.5.4. <http://www.gromacs.org>.
77. Parrinello, M.; Rahman, A., Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied Physics* **1981**, 52 (12), 7182-7190.
78. Bussi, G.; Donadio, D.; Parrinello, M., Canonical sampling through velocity rescaling. *The Journal of Chemical Physics* **2007**, 126 (1), 014101-7.
79. Lide, D. R., Handbook of Chemistry and Physics. **2008**, (88).
80. Back, P. J.; Woolf, L. A., (p,V,T,x) measurements for tetrahydrofuran and {xC<sub>4</sub>H<sub>8</sub>O + (1-x)H<sub>2</sub>O}. *Journal of Chemical Thermodynamics* **1998**, 30 (3), 353-364.
81. Chiang, J. Y., Bile acids: regulation of synthesis. *Journal of Lipid Research* **2009**, 50 (10), 1955-1966.
82. (a) Zhong, Z.; Yan, J.; Zhao, Y., Cholic Acid-Derived Facial Amphiphiles with Different Ionic Characteristics. *Langmuir* **2005**, 21 (14), 6235-6239; (b) Zhao, Y.; Zhong, Z., Oligomeric Cholates: Amphiphilic Foldamers with Nanometer-Sized Hydrophilic Cavities. *Journal of the American Chemical Society* **2005**, 127 (50), 17894-17901; (c) Luo, J.; Chen, Y.; Zhu, X. X., Invertible Amphiphilic Molecular Pockets Made of Cholic Acid. *Langmuir* **2009**, 25 (18), 10913-10917; (d) Lai, X.-Z.; Feng, Y.; Pollard, J.; Chin, J. N.; Rybak, M. J.; Bucki, R.; Epand, R. F.; Epand, R. M.; Savage, P. B., Ceragenins: Cholic

- Acid-Based Mimics of Antimicrobial Peptides. *Accounts of Chemical Research* **2008**, *41* (10), 1233-1240.
83. Brotherhood, P. R.; Davis, A. P., Steroid-based anion receptors and transporters. *Chemical Society Reviews* **2010**, *39* (10), 3633-3647.
84. Chhun, C.; Schmitzer, A. R., A pseudorotaxane umbrella thread with chloride transmembrane transport properties. *MedChemComm* **2011**, *2* (10), 987-990.
85. Chhun, C. Axe et Rotaxane Parapluie : vers de nouveaux transporteurs transmembranaires de chlorures et de médicaments cycliques. Université de Montréal, Montréal, 2011.
86. Eiben, A. E.; Smith, J. E., *Introduction to Evolutionary Computing*. Springer: 2003; p 300.
87. Blum, C.; Roli, A., Metaheuristics in combinatorial optimization: Overview and conceptual comparison. *ACM Computing Surveys* **2003**, *35* (3), 268-308.
88. Holland, J. H., *Adaptation in natural and artificial systems*. University of Michigan Press: Ann Arbor, 1975.
89. Fogel, D. B., What is evolutionary computation? *Spectrum, IEEE* **2000**, *37* (2), 26, 28-32.
90. Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E., The Protein Data Bank. *Nucleic Acids Research* **2000**, *28* (1), 235-242.
91. Huang, S.-Y.; Grinter, S. Z.; Zou, X., Scoring functions and their evaluation methods for protein-ligand docking: recent advances and future directions. *Physical Chemistry Chemical Physics* **2010**, (12), 12899-12908.

92. Meng, E. C.; Shoichet, B. K.; Kuntz, I. D., Automated docking with grid-based energy evaluation. *Journal of Computational Chemistry* **1992**, *13* (4), 505-524.
93. (a) Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J., Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *Journal of Computational Chemistry* **1999**, *19* (14), 1639-1662; (b) Huey, R.; Morris, G. M.; Olson, A. J.; Goodsell, D. S., A semiempirical free energy force field with charge-based desolvation. *Journal of Computational Chemistry* **2007**, *28*, 1145-1152.
94. Jones, G.; Willet, P.; Glen, R. C.; Leach, A. R.; Taylor, R., Development and validation of a genetic algorithm for flexible docking. *Journal of Molecular Biology* **1997**, *267* (3), 727-748.
95. Lorder, D. M.; Shoichet, B. K., Flexible ligand docking using conformational ensembles. *Protein Sci.* **1998**, *7* (4), 938-950.
96. Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S., Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *Journal of Medicinal Chemistry* **2004**, *47* (7), 1739-1749.
97. Corbeil, C. R.; Englebienne, P.; Yannopoulos, C. G.; Chan, L.; Das, S. K.; Bilimoria, D.; L'Heureux, L.; Moitessier, N., Docking Ligands into Flexible and Solvated Macromolecules. 2. Development and Application of Fitted 1.5 to the Virtual Screening of Potential HCV Polymerase Inhibitors. *Journal of Chemical Information and Modeling* **2008**, *48* (4), 902-909.
98. Thomas, P. D.; Dill, K. A., Statistical Potentials Extracted From Protein Structures: How Accurate Are They? *Journal of Molecular Biology* **1996**, *257* (2), 457-469.

99. (a) Huang, S.-Y.; Zou, X., An iterative knowledge-based scoring function to predict protein-ligand interactions: I. Derivation of interaction potentials. *Journal of Computational Chemistry* **2006**, (27), 1865-1875; (b) Huang, S.-Y.; Zou, X., An iterative knowledge-based scoring function to predict protein-ligand interactions: II. Validation of the scoring function. *Journal of Computational Chemistry* **2006**, (27), 1876-1882.
100. Yang, C. Y.; Wang, R.; Wang, S., M-score: a knowledge-based potential scoring function accounting for protein atom mobility. *Journal of Medicinal Chemistry* **2006**, 49 (20), 5903-5911.
101. Mooij, W. T.; Verdonk, M. L., General and targeted statistical potentials for protein-ligand interactions. *Proteins: Structure, Function, and Bioinformatics* **2005**, 61, 272-287.
102. Corbeil, C. R.; Moitessier, N., Docking Ligands into Flexible and Solvated Macromolecules. 3. Impact of Input Ligand Conformation, Protein Flexibility, and Water Molecules on the Accuracy of Docking Programs. *Journal of Chemical Information and Modeling* **2009**, 49 (4), 997-1009.
103. Moitessier, N.; Therrien, E.; Hanessian, S., A Method for Induced-Fit Docking, Scoring, and Ranking of Flexible Ligands. Application to Peptidic and Pseudopeptidic B-secretase (BACE 1) Inhibitors. *Journal of Medicinal Chemistry* **2006**, 49 (20), 5885-5894.
104. Böhm, H.-J., The development of a simple empirical scoring function to estimate the binding constant for a protein-ligand complex of known three-dimensional structure. *Journal of Computer-Aided Molecular Design* **1994**, (8), 243-256.
105. Åqvist, J.; Luzhkov, V. B.; Brandsdal, B. Ö., Ligand Binding Affinities from MD Simulations. *Accounts of Chemical Research* **2002**, 35 (6), 358-365.
106. Irwin, J. J.; Shoichet, B. K., ZINC: A Free Database of Commercially Available Compounds for Virtual Screening. *Journal of Chemical Information and Modeling* **2004**, 45 (1), 177-182.

107. Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J., Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Advanced Drug Delivery Reviews* **1997**, 23 (1-3), 3-25.